Structural and Chemical Investigation of Three Biosynthetic Enzymes

Ying Ge

This thesis is submitted in fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

January 2019

Candidate's declaration

I, Ying Ge, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 40,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2014.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date Signature of supervisors

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Ying Ge, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of 1 year on the following ground(s):

- Publication would preclude future publication

Supporting statement for printed embargo request

Some contents within the thesis are used in a manuscript which is being prepared for publication

Electronic copy

Embargo on all of electronic copy for a period of 1 year on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

Contents of the thesis are intended for use in future publications

Title and Abstract

- I require an embargo on the abstract only.

Date Signature of candidate

Date Signature of supervisors

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Ying Ge, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date Signature of candidate

Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a burgeoning family of natural products with desirable bioactivities. The cyanobactin class of RiPPs have demonstrated anticancer, anti-microbial activities and more, and have therefore attracted research attentions. Cyanobactins are synthesised as a precursor peptide (PatE homologues) containing recognition sequences and core peptides (sequences that become the final products), the latter of which undergo various modifications including heterocyclization, prenylation, oxidation and macrocyclization. The post-translational modification enzymes (PTMEs) utilised to modify the cyanobactin precursor have demonstrated magnificent versatilities that are potentially exploitable for the creation of libraries of natural and unnatural compounds. The structure and function of each of these enzymes are reviewed in Chapter 1, along with the current application of the cyanobactin biosynthetic pathway.

Chapter 2 describes my work on creating and characterising modified heterocyclases. The leader peptide was appended to heterocyclases that produce both thiazolines and (methyl)oxazolines. The modified catalysts are not only capable of utilising leaderless peptides, but also have higher yields than the native enzyme.

The next chapter investigates the reaction order of heterocyclase enzymes. The leader peptide was found to be partially responsible for the reaction order. In Chapter 4 the same investigation was applied to the oxidase, which aromatises and stabilises the heterocycles. The oxidase reaction was determined to be independent of the leader.

Chapter 5 delves into the phosphate chemistry of the heterocyclase, which uses ATP/Mg^{2+} via a kinase mechanism, but subsequently catalyses additional reactions that yield AMP and PP_i.

Finally, Chapter 6 describes the structural characterisation of *Psychrobacter arcticus* ATP phosphoribosyltransferase (ATPPRT), which is the first dedicated enzyme in histidine biosynthesis, and provides a key regulatory point for this pathway.

Contents

List of Figures

List of Tables

Abbreviations

ACN	Acetonitrile
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
AMP-CP	α,β-Methyleneadenosine 5-diphosphate
AMP-CPP	α,β-Methyleneadenosine 5-triphosphate
AMP-NPP	Adenosine 5'-(α,β-imido)triphosphate
AMP-PCP	β,γ-Methyleneadenosine 5-triphosphate
AMP-PNP	Adenosine 5'-(β,γ-imido)triphosphate
ATP	Adenosine triphosphate
ATP-PRT	ATP phosphoribosyltransferase
BGC	Biosynthetic gene clusters
BME	β-mercaptoethanol
CCP4	Collaborative Computing Project Number 4
CTD	Carboxyl-terminal domain
DMAPP	Dimethylallyl pyrophosphate
DMSPP	Dimethylallyl thiolodiphosphate
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease I
DTT	Dithiothreitol
DUF	Domain of unknown function
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionisation
FA	Formic acid
FMN	Flavin mononucleotide
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
HTH	Helix-turn-helix
IDA	Information-dependent acquisition
IPTG	Isopropylthio-β-D-galactoside
LAPs	Linear azol(in)e-containing peptides
LB	Luria-Bertani (growth medium)
LC	Liquid chromatography
MALDI	Matrix-assisted laser desorption ionisation
MALS	Multiangle Light Scattering

MESG	2-amino-6-mercapto-7-methylpurine riboside
MS	Mass spectrometry
MSD	Mass selective detector
MSMS	Tandem mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut off
m/z	Mass over charge
NMR	Nuclear magnetic resonance
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase
NTA	Nitrilotriacetic acid
NTD	Amino-terminal domain
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PCP	Peptidyl carrier protein
P_i	Inorganic phosphate
PIS	Product ion scan
PKS	Polyketide synthetase
PNP	Purine nucleoside phosphorylase
PP_i	Inorganic Pyrophosphate
PPase	Pyrophsphatase
PRATP	Phosphoribosyl-ATP
PRPP	Phosphoribosyl pyrophosphate
PTME	Post-translational modification enzyme
RiPP	Ribosomally synthesized and post-translationally modified peptide
RNA	Ribonucleic acid
RRE	RiPP recognition element
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SUMO	Small ubiquitin-like modifier
TCEP	Tris(2-carboxyethyl)phosphine
TE	Thioester
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid
TIC	Total ion count
TLS	Translation, libration and screw-rotation
TOF	Time of flight
TOMMS	Thiazole/oxazole-modified microcins
Tris	Trisaminomethane
UV	Ultra-violet

Standard Amino Acids

Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparginine	Asn	N	Methionine	Met	M
Aspartic Acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamic Acid	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucin	Ile	I	Valine	Val	V

Heterocyclized Amino Acids

Thiazoline	C^{het}	Thn	Oxazole	$S^{het\text{-}ox}$	Oxz
Thiazole	$C^{het\text{-}ox}$	Thz	Methyloxazoline	T^{het}	MOxn
Oxazoline	S^{het}	Oxn	Methyloxazole	$T^{het\text{-}ox}$	MOxz

Nucleic Acids

Adenosine A Cytosine C Guanine G Thymidine T

List of Enzymes

MicD	Heterocyclase from *Microcystis aeruginosa*
LynD	Heterocyclase from *Lyngba* spp.
PatD	Heterocyclase from *Prochloron didemni*
OscD	Heterocyclase from *Oscillatoria nigro-viridis PCC 7112*
TenD	Heterocyclase from *Nostoc spongiaeforme* var. *tenue*
ArtGox	Azoline oxidase from from *Arthrospira platensis*
ThcOx	Azoline oxidase from *Cyanothece sp. PCC 7425*
*Pa*ATPPRT	ATP phosphoribosyltransferase from *Psychrobacter arcticus*
*Ll*ATPPRT	ATP phosphoribosyltransferase from *Lactoccocus Lactis*
*Tm*ATPPRT	ATP phosphoribosyltransferase from *Thermotoga maritima*
HisG	Catalytic subunit of ATPPRT
HisZ	Regulatory subunit of ATPPRT
*Mt*PPase	Pyrophosphatase from *Mycobacterium tuberculosis*

Acknowledgements

First of all, I would like to express my sincerest gratitude to my supervisor, Prof Jim Naismith, for providing the opportunity for me to conduct research in his group, guidance and support throughout my graduate studies. I also owe my thanks to Dr Petr Killian, who took over the administrative sides of my supervision after Jim relocated to Oxford, and Prof Terry Smith, my internal examiner, who has offered many useful advices over the course of my PhD.

I am grateful to all members of the JHN group, in particular, Dr Clarissa Melo Czekster, Dr Huanting Liu, Dr Magnus Alphey, Hannes Ludewig and Sisi Gao. I am also very thankful to many academic and administrative staff from the BSRC and School of Chemistry the following members of the BSRC, in particular, those who provided expertise in techniques crucial to my research: Dr Catherine Botting, Dr Sally Shirran, Dr Silvia Synowsky and Dr Andrew Bent for mass spectrometry; Dr Uli Schwarz-Linek and Dr Ona Miller for NMR.

I enjoyed and benefited tremendously from the ATP-PRT project, a collaboration with Dr Rafael da Silva's group. I would like to express my great appreciation to Dr Magnus Alphey for his assistance in crystallography with regard to this project and in general, and Dr Garib Murshudov (University of Cambridge) for helping with the structure.

I would like to acknowledge the generous donations of the vector pBMS by Dr Huanting Liu, pJ411-PPA by Dr Luiz de Carvalho (The Francis Crick Institute) and the peptide ITACITACAYDGE with C-terminal amide by Dr Brunello Nardone.

Last but not least, I owe my gratitudes to my friends, family, and my cat Meowthew, for supporting me through the toughness of postgraduate studentship. In particular, I thank Hannes Ludewig for proofreading this thesis.

I dedicate this thesis to my grandfather who sadly passed away this June. He was a promising student who could not go to university due to economic hardship but had since supported his children and grandchildren through their education.

Funding

The work is supported by ERC grant NCB-TNT (339367).

Publications

- Czekster, C. M., **Ge, Y.** and Naismith, J. H. (2016) Mechanisms of cyanobactin biosynthesis. *Curr Opin Chem Biol* 35, 80-88. (Joint first author)
- Stroek, R., **Ge, Y.**, Talbot, P.D., Glok, M.K., Bernas, K.E., Thomson, C.M., Gould, E.R., Alphey, M.S., Liu, H., Florence, G. J., *et al.* (2017) Kinetics and Structure of a Cold-Adapted Hetero-Octameric ATP Phosphoribosyltransferase. *Biochemistry* 56, 793-803. (Joint first author)
- Alphey, M. S., Fisher, G., **Ge, Y.**, Gould, E. R., Machado, T. F. G., Liu, H., Florence, G. J., Naismith, J. H. and da Silva, R. G. (2018) Catalytic and Anticatalytic Snapshots of a Short-Form ATP Phosphoribosyltransferase. *ACS Catalysis* 8, 5601-5610.
- Gao, S., **Ge, Y.**, Bent, A. F., Schwarz-Linek, U. and Naismith, J. H. (2018) Oxidation of cyanobactin precursor peptide is independent of leader peptide and operates in a defined order. *Biochemistry* 57, 5996-6002. (Joint first author)

Chapter 1

Introduction

1.1 Peptide Natural Products

Bioactive peptides produced by organisms such as bacteria and fungi provide a plethora of substances from which numerous drug leads have been, and continue to be identified [1–3]. The majority of them yet unexplored, this vast family of compounds have already yielded widely used drugs, including vancomycin and insulin. More recently, novel antibiotics teixobactin [4] and malacidins [5] were discovered, both of which are peptide-based natural products. With modern medicine being threatened by the rise of antibiotic resistance and drug tolerance, it is imperative that we inquire into the solutions nature has to offer, notably, the rich repertoire of peptide-derived secondary metabolites. Bioactivity of peptide natural products comes from their diverse structures and lush reserve of non-proteinogenic moieties, which in turn originate from the way they are biosynthesised. In general, these peptides can be divided into two categories based upon their biosynthesis: nonribosomal peptides (NRPs), like teixobactin and malacidins, and ribosomally synthesized and post-translationally modified peptides (RiPPs). Due to the structural and chemical complexity of both NRPs and RiPPs, total- and semi-synthesis of these face major challenges, and their biosynthetic mechanisms have been under extensive investigation as a result.

1.1.1 Nonribosomal Peptides

Secondary metabolites of this class include the antibiotic vancomycin, the anticancer agent bleomycin and the immunosuppressant cyclosporin. Nonribosomal peptides are synthesised by multimodular nonribosomal peptide synthetases (NRPS), which work independently of mRNAs and the ribosomal machinery, but instead use bespoke building blocks to produce highly specific peptides [6–8]. Diverse building blocks are made from

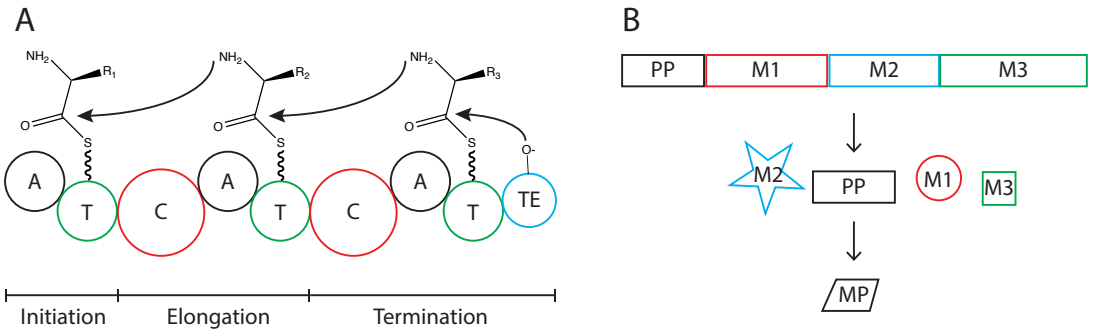

Figure 1.1: Overview of NRP and RiPP biosyntheses. (A) schematic showing the general assembly line of nonribosmal peptide synthetases. Round shapes labelled 'A', 'T', 'C' and 'TE' represent the adenylation, thiolation, condensation and thioesterase domains, respectively. (B) schematic of RiPP synthesis. PP, precursor peptide; M1, M2 and M3, modifying enzymes; MP, mature peptide.

proteinogenic amino acids and other substrates before they are assembled by NRPS into a polypeptide chain, which can then be further modified and decorated [7]. NRPS are closely related to polyketide synthetases (PKS) and the two are often hybridised [9], further increasing product diversity. Characterisation of NRPS mainly focused on bacterial and fungal synthetases, but these synthetases have also been found in higher organisms like drosophila [10] and nematodes [11]. The overall architecture of NRPS is that of an assembly line (Figure 1.1A) [7], with an initiation module, one or more elongation module(s), and a termination module, each of which in turn consists of multiple domains. The initiation modules contains an adenylation domain (A), which selects and activates amino acids using ATP, turning them into a high-energy aminoacyl-AMP species, and a thiolation domain (T), also known as a peptidyl carrier protein domain (PCP), which subsequently takes up the aminoacyl group with its -SH moiety. Following the T domain is the condensation domain (C) of the first elongation module, which is succeeded by its own set of A and T domains. Catalysed by the C domain, an amide bond is formed between the first and the second amino acid, displacing the thioester bond that tethers the first amino acid to its carrier protein. The same process occurs between the first and the second elongation module, and so forth. After the last transfer, an thioesterase domain (TE) from the termination module releases the full-length polypeptide, and in some cases cyclises it [6]. Additional domains may be present in each module, for modifications of the nascent peptide such as epimerisation [12], methylation [13] and oxidation [14]; alternatively, *trans*-acting enzymes may be recruited for post-NRPS decorations [7]. As a general rule, each building block requires its own adenylation and thiolation proteins, and a condensation domain (or a variant) is required between every two modules. NRPSs are typically very large as a result of this, exemplified by the 1.4 MDa cyclosporin synthetase [15].

1.1.2 Ribosomally Synthesised and Post-Translationally Modified Peptides

Ribosomally synthesized and post-translationally modified peptides [16] represent another expanding class of natural products that have exhibited various bioactivities such as anti-viral and anti-fungal [17, 18]. RiPP synthesis (Figure 1.1B) starts with a precursor peptide made by the ribosome using the 20 common amino acids, which comprises a leader peptide, further recognition sequences and a core sequence (the part that yields the final product). The precursor is modified by a series of post-translational modification enzymes (PTMEs) during maturation. These modifications enrich the peptides with chemical groups and intramolecular structures, giving them complexities that rival NRPs. Substantial work, led by the van der Donk group, has been carried out to characterise lanthipeptides, compounds that comprises the lanthionine moiety (Figure 1.2A) [19], whilst linear azol(in)e-containing peptides (LAPs, Figure 1.2B), cyanobactins (Figure 1.3) and thiopeptides (Figure 1.2C) are among the other families of best studied RiPPs [16, 20–22]. Unlike NRP synthesis, RiPP synthesis circumvented the need for elaborate multidomain synthetases, which can be burdensome for microorganisms and also difficult to reconstitute and engineer *in-vitro* [23–25]. This however also meant their biosynthetic gene clusters (BGCs) are less conspicuous than NRPS genes, which has to encode, in multiplicity, essential modules of the assembly line (Figure 1.1A). It was the development of genome mining tools that drastically expanded the number of RiPPs, and prompted scientists to unify their classification and nomenclature [16, 26], even though sporadic discoveries of the RiPP biosynthetic clusters date back to the 20th century [27]. To date, tens of subfamilies of RiPPs have been characterised including the above mentioned families, as well as bottromycins, lasso peptides and proteusins [16, 28]. RiPP biosynthesis and the function of various PTMEs have been of great interest in recent years, largely due to the fact that recognition between the precursor and the PTMEs are usually mediated by conserved elements outside the core peptide, whereas the core itself is hypervariable [16]. The uncoupling of recognition-catalysis not only enabled these enzymes to process multiple natural substrates and those from closely related pathways, but also opened up horizons for hybridisation, splicing and engineering, where the recognition elements on substrates can be altered to lead them to (multiple) destined enzymes [29, 30]. Moreover, core sequences can also be readily edited by means of molecular biology, exponentially increasing substrate varieties. Together, the promiscuity of PTMEs and the alterability of precursor sequences make the RiPP machinery a promising tool for the biosynthesis of diverse peptide derivatives.

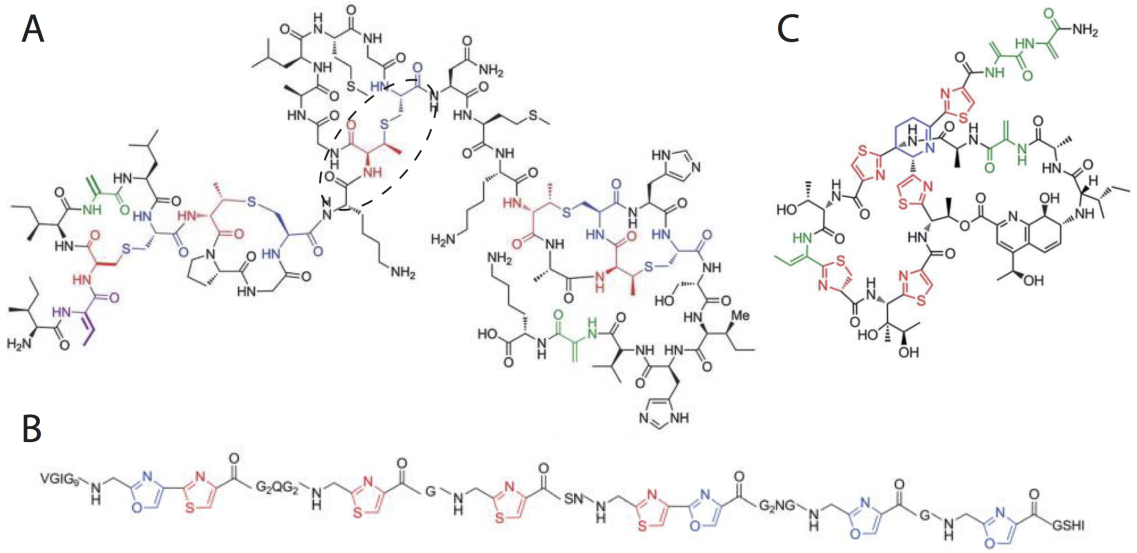

Figure 1.2: Examples of RiPPs. (A) Nisin A, an antibiotic lanthipeptide (lantibiotic) commonly used as a food preservative. Dashed circle denotes the lanthionine moiety that gives the name lantipeptides. (B) Microcin B17, an antibiotic LAP. (C) Thiostrepton A, a thiopeptide containing four thiazoles and a thiazoline. Figures taken and adapted from [16].

1.1.3 The Cyanobactin Family of RiPPs

Among the best characterised RiPPs are cyanobactins, such as patellamides, trunkamides and ulicyclamides. Drug-like properties, including cytotoxicity, anti-tumour and anti-malarial activities have been reported for this class of natural products [31–35], making them interesting subjects of biomedical studies. First isolated from the marine ascidian *Lissoclinum patella* in the 1980s [36], cyanobactins were later attributed to *L. patella*'s cyanobacterial symbiont *Prochloron didemni* [37, 38]. Discoveries of cyanobactins from other species were made throughout the 1990s and 2000s; in addition, compounds have been identified after predictions made by means of metagenomic mining, resulting in a great expansion of the size and diversity of the cyanobactin family [20, 21, 39, 40]. The archetypical cyanobactin is an azol(in)e-containing head-to-tail macrocycle that is 6-11 amino acids in length [21], but members of the group also include homodetic macrocycles that can be over 20 amino acids in length, as well as short, linear peptides featuring extensive post-translational modifications such as *N*-prenylation and *O*-methylation [40]. An example of each type of cyanobactins is shown in Figure 1.3. Macrocyclic cyanobactins have been of particular interest to researchers due to their rigid conformation that facilitates binding to specific targets and increases resistance to degradation [41]. Chemical synthesis of macrocycles usually involves complicated multiple-step processes (Figure 1.4) [42] that are not readily applicable to a wide range of derivatives [42, 43], making the biosynthesis of them an attractive alternative.

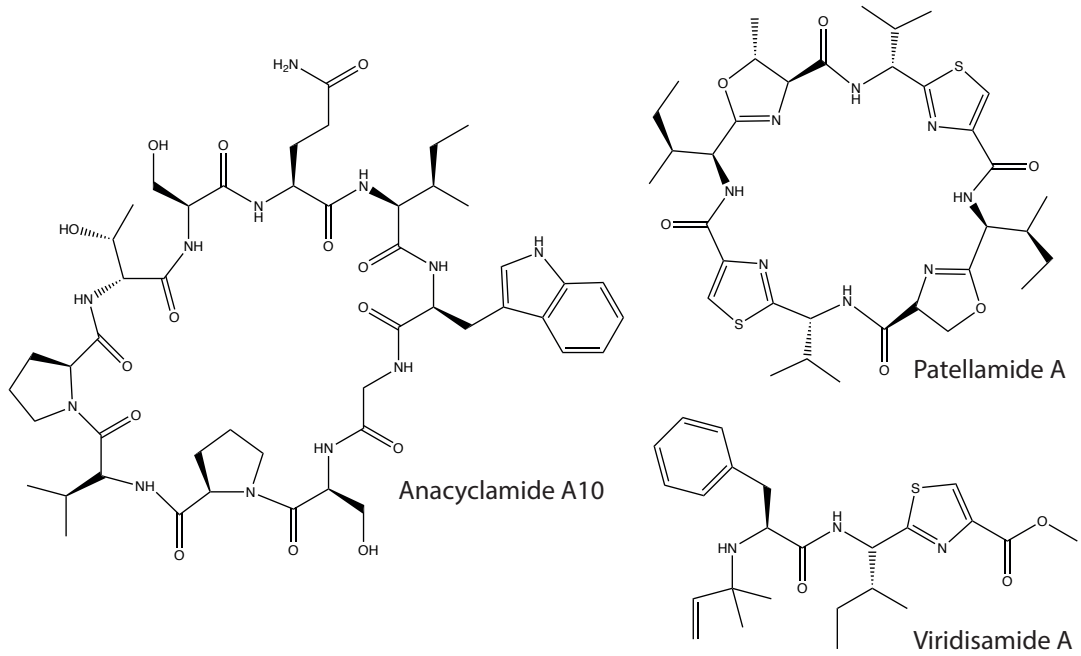

Figure 1.3: Examples of Cyanobactins. Patellamide A, a heterocycle-containing macrocycle; Viridisamide A, a short, linear cyanobactin peptide; Anacyclamide A10, a homodetic macrocyclic compound.

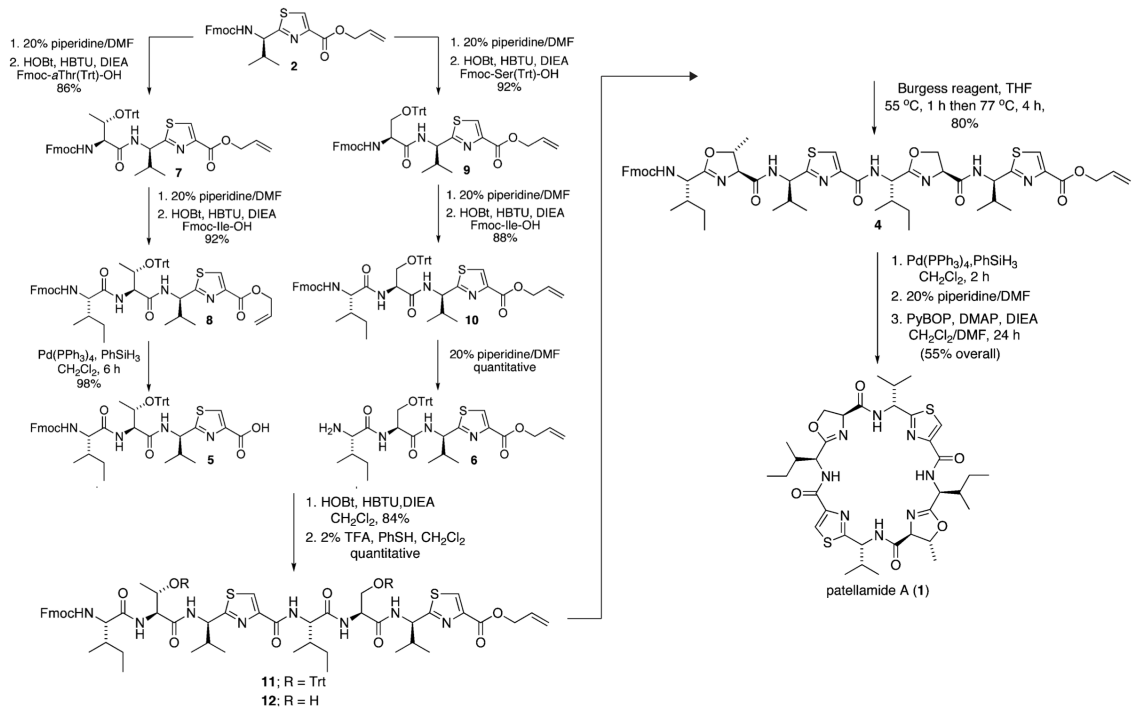

Figure 1.4: Total synthesis of Patellamide A as reported by Garcia-Reynaga and colleagues [42]. Figure adapted from [42].

1.2 Overview of Cyanobactin Biosynthesis

The biosynthetic machinery of cyanobactins contains multiple PTMEs encoded in a single gene cluster, along with the precursor peptide. Patellamides are the first of cyanobacterial peptides to have the whole biosynthetic pathway reconstituted and characterised [37, 38], and will be used as the model system here. The *pat* gene cluster encodes seven genes, patA-patG, corresponding to the precursor peptide (PatE) and its PTMEs (Figure 1.5A). The canonical PatE is composed of 71 amino acids: an N-terminal 37-residue leader sequence, followed by two 8-residue core peptide cassettes each flanked by a protease recognition sequence G(L/V)E(A/P)S on the N-terminus and a macrocycliztion signature AYD on the C-terminus (Figure 1.5B). The cassettes shown in Figure 1.5B, VTACITFC and ITVCISVC, encode patellamide C (Figure 1.5C) and patellamide A (Figure 1.3), respectively. The proteins encoded with it are: PatA, a protease plus a domain of unknown function (DUF); PatB and PatC, proteins of unknown function; PatD, a heterocyclase (cyclodehydratase); PatF, a putative prenylase and PatG, a tripartite protein with an oxidase domain, a protease (macrocyclase) domain and a DUF. With the exception of PatB and PatC, which although conserved are non-essential to patellamide biosynthesis [20, 21], the structures and functions of each of these PTMEs are reviewed in details in the following sections. Figure 1.5C gives a chemist's view of patellamide C with annotations showing sites of modifications by each enzyme.

1.3 Heterocyclase

Heterocyclization (cyclodehydration) of serines, threonines and cysteines by PatD is considered to be the first step in Patellamide biosynthesis [16, 44, 45]. In addition to patellamides, azol(in)e-containing cyanobactins included macrocyclic ones such as trunkamides [46, 47] and microcyclamides [35, 48, 49], as well as linear ones like viridisamide A [40]. Phylogenetic studies [40, 50, 51] showed that clusters capable of biosynthesizing cyanobactins that bear heterocycles encode a YcaO domain-containing heterocyclase (cyclodehydratase), and optionally, an FMN-dependent oxidase (discussed later in section 1.5). These genes are missing from biosynthetic clusters with products that do not contain heterocycles. This section explains the roles of heterocycles in natural products, reviews investigations of the cyanobactin heterocyclase and related enzymes, and compares its heterocyclization mechanism to that employed by NRPSs.

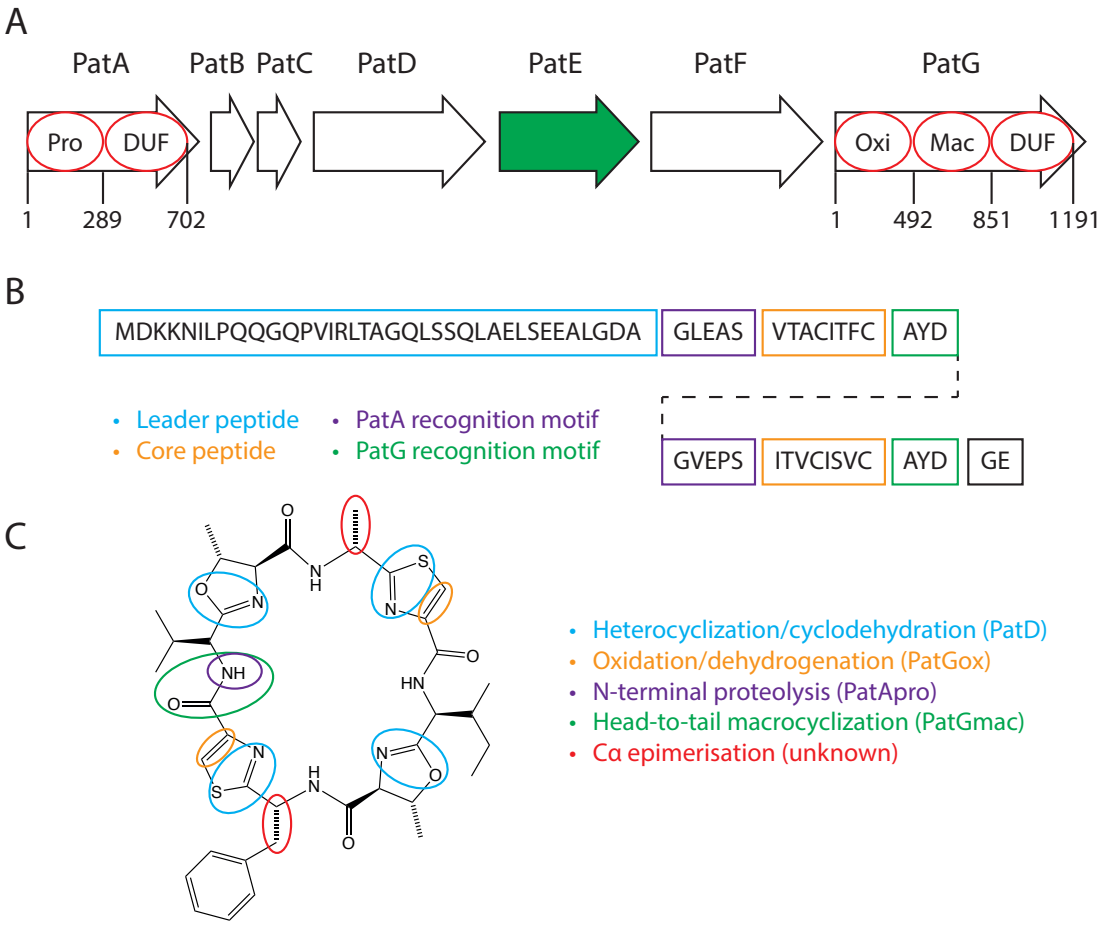

Figure 1.5: The Cyanobactin biosynthetic pathway. (A) layout of the Patellamide biosynthetic gene cluster. Pro, protease; DUF, domain of unknown function; Oxi, oxidase; Mac, macrocyclase. (B) Annotated sequence of the precursor peptide PatE containing two cassettes - VTACITFC (patellamide C) and ITVCISVC (patellamide A). (C) A chemist's view of Patellamide C.

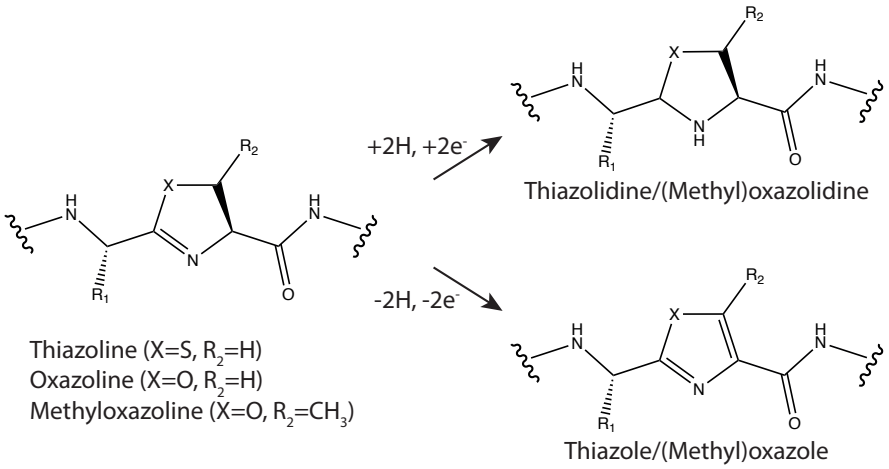

Figure 1.6: Azolines, Azolidines and Azoles shown in a peptidyl context. X=S, O; R_2=H, CH_3.

1.3.1 Azol(in)es in Natural Products

The heterocycles found in cyanobactins are five-membered thiazolines, oxazolines and methyloxazolines (Figure 1.6), and their oxidised versions thiazoles and (methyl)oxazoles [40, 50, 51]. The reduced forms thiazolidines and (methyl)oxazolidines (Figure 1.6) also exist in natural products [52, 53], although not in cyanobactins. Azolines usually arise in a dipeptide context, where the side chain of serine, threonine or cysteins cyclises onto its preceding carbonyl group [52], but non-peptidyl moieties such as hydroxyphenyl groups have been seen to precede oxazolines in compounds such as bacterial siderophores [54–56]. Tandem arrays of heterocycles are not uncommon [52], and they may undergo further modifications such as halogenation [57] and methylation [58]. Heterocycles reduce polarity, add chemical diversity and conformational rigidity to peptides, as well as provide recognition elements that interacts with specific targets [52, 59]. They are shown to chelate iron in the siderophores mycobactin [54], vibriobactin [56] and yersiniabactin [53], and Zinc/Magnesium in the antibiotic bacitracin [60], to act as DNA intercalators in the anticancer agent bleomycin A2 [61], and there is evidence suggesting that they form sequence-dependent contacts with RNAs and proteins [62, 63]. Among RiPPs azol(in)es are prevalent in LAPs, thiopeptides and cyanobactins, compounds with varied bioactivities such as cytotoxicity [16], to which heterocyclization may have contributed in different ways.

1.3.2 Chemistry of Heterocyclization

The heterocyclization reaction in cyanobactin biosynthesis is both ATP and magnesium dependent [64, 65], as are other reactions catalysed by YcaO domain-containing enzymes such as the thiazole/oxazole-modified microcin (TOMM) heterocyclase BalhD from *Bacillus sp.* Al Hakam [66–69]. Formation of heterocycles in RiPPs starts with the

hydroxyl or sulfhydryl group from a serine, threonine or cysteine side chain attacking the amide bond N-terminal to it [70], leading to the formation of a tetrahedral intermediate called hemiorthoamide (Figure 1.7A showing cysteine residue thiol as a nucleophile). ATP hydrolysis was proposed to power motions in the enzyme, which functions as a molecular machine of some sort [52, 66], but stoichiometric analysis pointed to a mechanism where ATP is directly involved in the chemical reaction of heterocyclization [71]. According to this mechanism, the activated tetrahedral oxygen attacks ATP on one of the phosphorus atoms, forming a phosphorylated or pyrophosphorylated intermediate and releases ADP/AMP; the adduct is subsequently eliminated to give thiazoline/oxazoline [71]. An unambiguous mechanism for this step is yet to be established, with various experimental evidence supporting both possibilities [45]. ADP and inorganic phosphate (P_i) were observed as the product of ATP hydrolysis during microcin B17 biosynthesis and in the BalhC/D systems, pointing to the enzyme being a kinase (Figure 1.7A) [70, 71]. Structural analyses of ATP-bound *E. coli* YcaO and LynD supported this, showing the γ-phosphate exposed to attack and the β-phosphate shielded by the protein [64, 71]. However, AMP and pyrophosphate (PP_i) have been detected during catalysis by TruD and LynD [64, 65], contradicting the kinase mechanism. Furthermore, *E. coli* YcaO, the function of which is unknown, hydrolyses ATP to AMP and PP_i [72].

During non-ribosomal peptide synthesis, on the other hand, ATP/Mg^{2+}-independent heterocyclization is carried out by the Cy domain, a variant of the canonical condensation domain [73]. Structurally the Cy domain resembles the C domain, although the highly conserved active site HHxxxDG motif in C is replaced with a DxxxxD motif [74–76]. Nevertheless, Cy still performs condensation reactions (as well as cyclodehydration), and the DxxxxD motif is believed to play a structural rather than a catalytic role in these proteins [73–75]. Mutational analyses of the bacillamide synthetase Cy domain (BmdB Cy2) showed that residues D1226 and T1196 are essential for heterocyclization, but not condensation [74]. These residues are 2.9 Å apart in the BmdB Cy2 structure [74], and conserved (2.7 Å apart) in the epothilone synthetase Cy domain (EpoB Cy) [75], making them plausible candidates for catalytic residues. Figure 1.7B shows the mechanism of heterocyclization proposed for BmdB Cy2, which occurs after the condensation of cysteine and its preceding residue [74].

1.3.3 Structure of the Cyanobactin Heterocyclase

Structures of the heterocyclases TruD (*P. didemni*) and LynD (*Lyngbya* spp.) were solved by X-ray crystallography [64, 65]. TruD and LynD share 79% sequence identity with each other and are almost identical structurally (r.m.d.s of 1.30 over 779 residues). Figure 1.8 shows the structure of LynD in complex with PatE (partially visible) and the non-hydrolysable ATP analogue AMP-PNP (β,γ-Imidoadenosine 5′-triphosphate). The

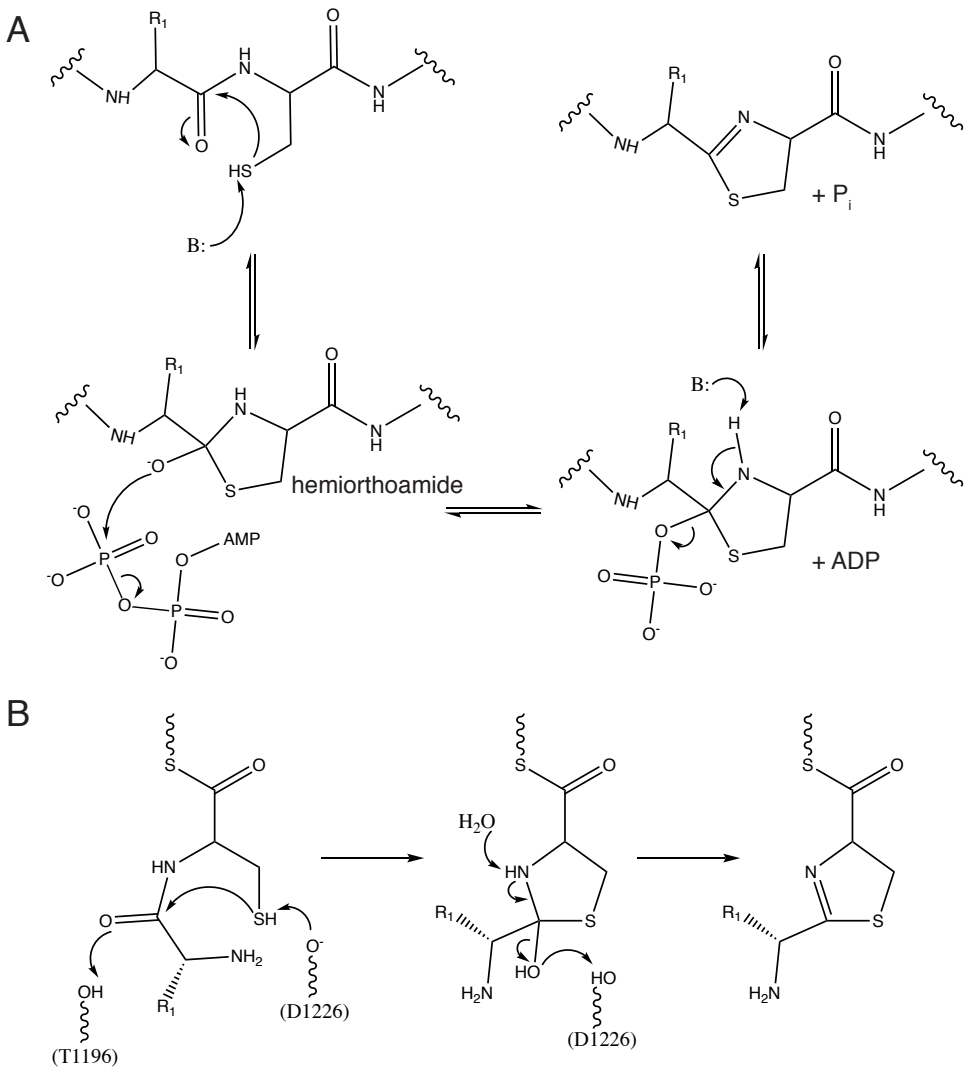

Figure 1.7: Heterocyclization mechanisms proposed for RiPPs and NRPS. (A) Kinase model of heterocyclization by YcaO-domain heterocyclases. (B) Proposed mechanism of heterocyclization by the Cy domain of the NRPS BmdB.

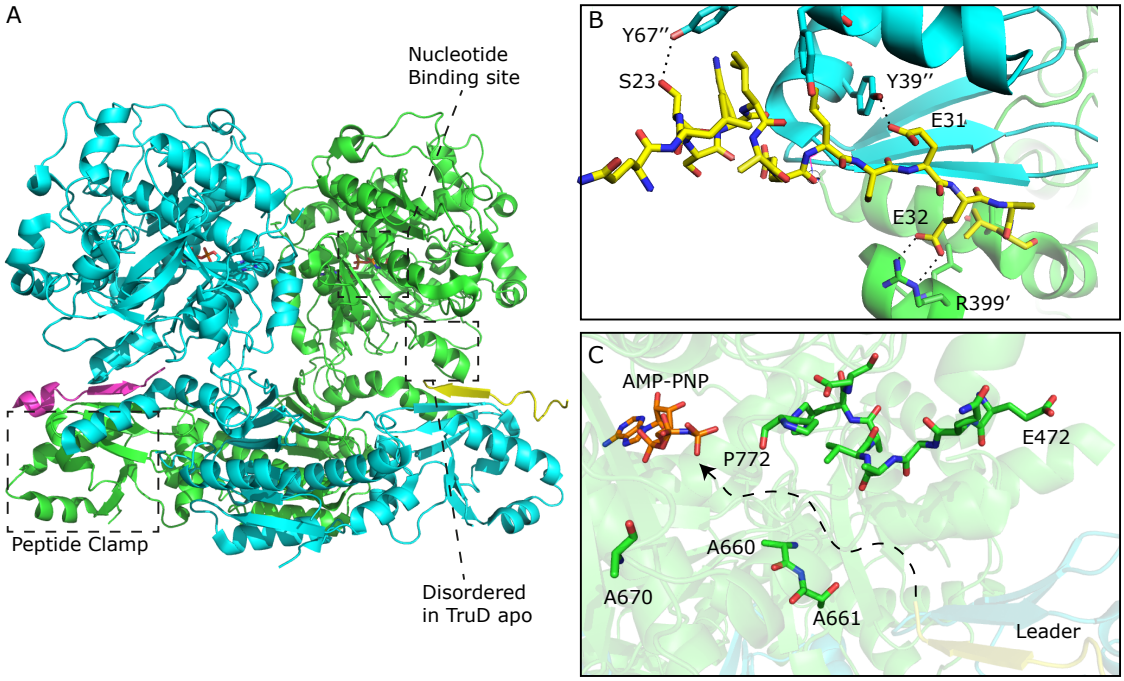

Figure 1.8: Crystal structure of LynD in complex with substrates. (A) Overview of the LynD dimer. The two chains are in green and cyan; PatE (leader) sequences are shown in yellow and magenta; AMP-PNP is shown as sticks. The peptide clamp, nucleotide-binding site and the region ordered by leader binding are highlighted by dashed boxes. (B) Close-up of the interface between LynD and PatE leader. Colour schemes are as in (A). Residues from the leader and key residues from LynD are presented by sticks; dashed lines signify hydrogen bonds; residues from the green chain is marked with a single prime; residues from the cyan chain are marked with a double prime. (C) Key residues that differ between LynD and heterocyclases whose products contain oxazolines.

molecule is oriented in such a way that the C-terminal domain (CTD) is on the top and the N-terminal domains (NTDs) sit at the bottom. Domain 1 and 2 (NTDs) of TruD/LynD (residues 1-315 of LynD) share limited sequence homology with the adenylating enzyme MccB (an E1 ubiquitin-activating enzyme), with two conserved cysteines in a 'CXXC' motif coordinating a Zinc ion; However, residues of the nucleotide binding sites in the MccB domain are not well conserved in LynD or TruD, making an MccB-like adenylating mechanism unlikely. The function of the MccB domain in LynD is proposed to instead lie in the 'peptide clamp' motif (Figure 1.8), consisting of three alpha helices and three beta sheets, otherwise known as a winged helix-turn helix (HTH) motif. The active site of the heterocyclase locates in the C-terminal, YcaO-like domain (residues 316-775), which is shown to bind AMP, ADP + P_i and AMP-PNP in structures of LynD complexed with these nucleotides [64]. However, attempts to obtain density for the the core sequence of PatE, which is modified by LynD, has been unsuccessful. Unlike the conversed leader peptide which makes specific contacts with LynD, the core region is hypervariable and probably has a higher level of a flexibility inside the active site.

1.3.4 Substrate Recognition by the Heterocyclase

PatD-like enzymes have demonstrated very limited selectivity towards residues flanking the transformed residue within core region, but their recognition of substrates depends a conserved sequence within the N-terminal leader peptide [64, 65, 77]. This strategy, also employed by a wide range of other PTMEs [16, 78], allows them to synthesise highly variable products with the same catalyst. Despite the lack of sequence similarities between the leader peptides of different classes of RiPPs and their cognate PTMEs, there is a shared structural motif (the 3-helices, 3-sheets peptide clamp, as described in the last paragraph) involved in protein-peptide recognition, which was discovered by a bioinformatics study [79] and designated RiPP precursor peptide recognition element (RRE) - this designation is used interchangeably with 'HTH motif, 'peptide clamp' and 'RiPP domain' in this thesis. For instance, LynD and the lantibiotic nisin dehydratase NisB recognise residues within and surrounding the consensus sequences 'LAELSEEAL' (PatE residues 26 to 34) and 'FNDL' (NisA numberings -18 to -15), respectively [64, 80], but both enzymes utilise structurally similar recognition domains (Figure 1.9). In both cases, the leader forms a range of specific contacts with the HTH motif, including hydrophobic interactions, hydrogen bonds and salt bridges, and adds a fourth beta sheet to it [64, 80]. In LynD, leader binding also regularises a flexible region (residues 371-415) of the YcaO domain (Figure 1.8), which was disordered in the apo structure of TruD, and turns it into part of the active site [64, 65]. These interactions activate the enzyme and likely also orient the rest of the peptide in such a way that the core sequence reaches the active site. Unfortunately, the core peptide was seen in neither the LynD [64] nor the NisB structure [80], making it difficult to elaborate on the role of leader in orienting the peptides.

In contrast, how the leader activates the cyanobactin heterocyclase is much better characterised. Activation happens both when the leader is *in cis* with the core, like in the native substrate, and when the leader and the core are added *in trans* [64, 77]. Cognate leader peptides added *in trans* have also been shown to enhance the activities of other PTMEs, such as the lacticin 481 synthetase LctM and the nisin synthetases NisB and NisC [81, 82]. In addition, there are a number of cases where constitutively active versions of PTMEs were created by fusing their cognate leader peptides to them [29, 64, 83]. The observation of *trans*-activation suggests that these enzymes have affinity for the core peptides themselves, as in these cases binding of leader does not bring the core peptides into spatial proximity with the enzymes. This native affinity for the core peptide can however be enhanced by leader binding, as is the case between the lantibiotic synthetase HalM2 and its cognate peptide [29]. It should be pointed out that the aforementioned *trans*-activated and fused enzymes vary in their efficacy to process their substrates fully; in other words, having the leader peptides *in cis* brings additional benefits. Leader binding could contribute to the the correct positioning of core peptides, despite a lack of direct evidence.

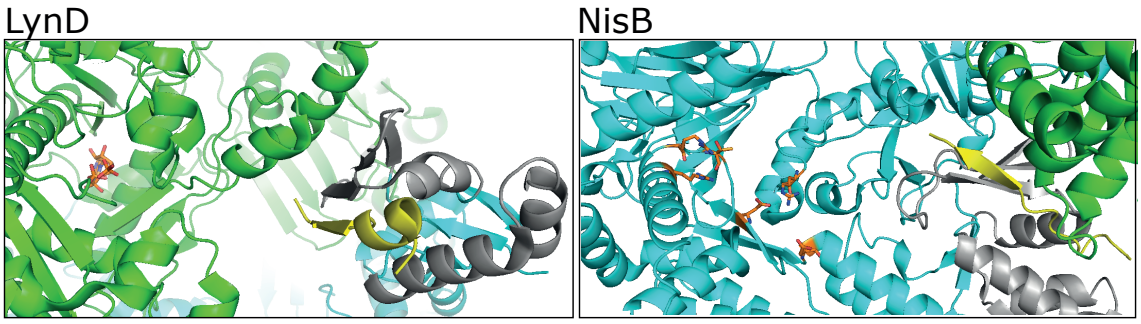

Figure 1.9: Structural resemblance of leader peptide binding in LynD and NisB. LynD and NisB protein chains are shown in green and cyan, except for the peptide clamp domain which is coloured grey; the leader peptide is shown in yellow; AMP-PNP in LynD and active site residues in NisB are shown by stick presentation. The leader peptide adds a beta-strand to the peptide clamp motif.

1.3.5 Specificity and Directionality of the Heterocyclase

Three amino acids, cysteine, threonine and serine, are substrates of the heterocyclase. Depending on the enzyme, however, not all heterocyclizable residues are converted into azolines. There are mainly two groups of enzymes, one of which, like PatD, catalyses both cysteines and threonines/serines [67], whereas the other group, exemplified by TruD, only produces thiazolines [65]. No enzyme has been reported to make oxazolines exclusively, likely due to the higher nucleophilicity of the thiolate anion (Cys S^-) compared to the alkoxide (Ser O^-) (pKa 8 vs 13). In addition, location of the residue within the core peptide may also determine whether it is heterocyclization, for example, the cyanobactin Comoramide A contains two threonines in the same cassette, of which only one is heterocyclized [86]. It is not clear what causes the chemo- and regio-selectivity of these enzymes. PatD shares 88% sequence identity with the cysteine-only catalyst TruD, but is only 73% and 77% identical, respectively, with the enzymes MicD (*Microcystis aeruginosa*) and TenD (*Nostoc spongiaeforme* var. *tenue*), which have been shown to make oxazol(in)es [87]. A closer look however reveals that the numbers are skewed by the almost identical (99% of sequence) N-terminal domains of PatD and TruD, and when YcaO domains (residues 328-781 as of TruD) alone are compared, PatD were shown to be more closely related with MicD and TenD than it is with TruD (Figure 1.10, Figure 1.11). The sequences are still very similar globally (Figure 1.10), and only a few segments that were common in PatD, TenD and MicD are not shared by the enzymes with thiazol(in)e-only products (TruD, LynD, OscD): an 'AE/QGSxAAYR' sequence that corresponds to residues 479-484 in TruD (472-479 in LynD), a 'TV' motif (TruD K620-A621, LynD K613-T614), an 'SK' motif (TruD A667-T668, LynD A660-D662), a glutamate (TruD A677, LynD A670), a proline (TruD A686, LynD T679) and a asparagine (TruD P778, LynD P772). Equivalent residues in the LynD structure (Figure 1.8C) are found mostly on the the solvent-accessible surface of the protein. Residues 472-479 are located in a loop,

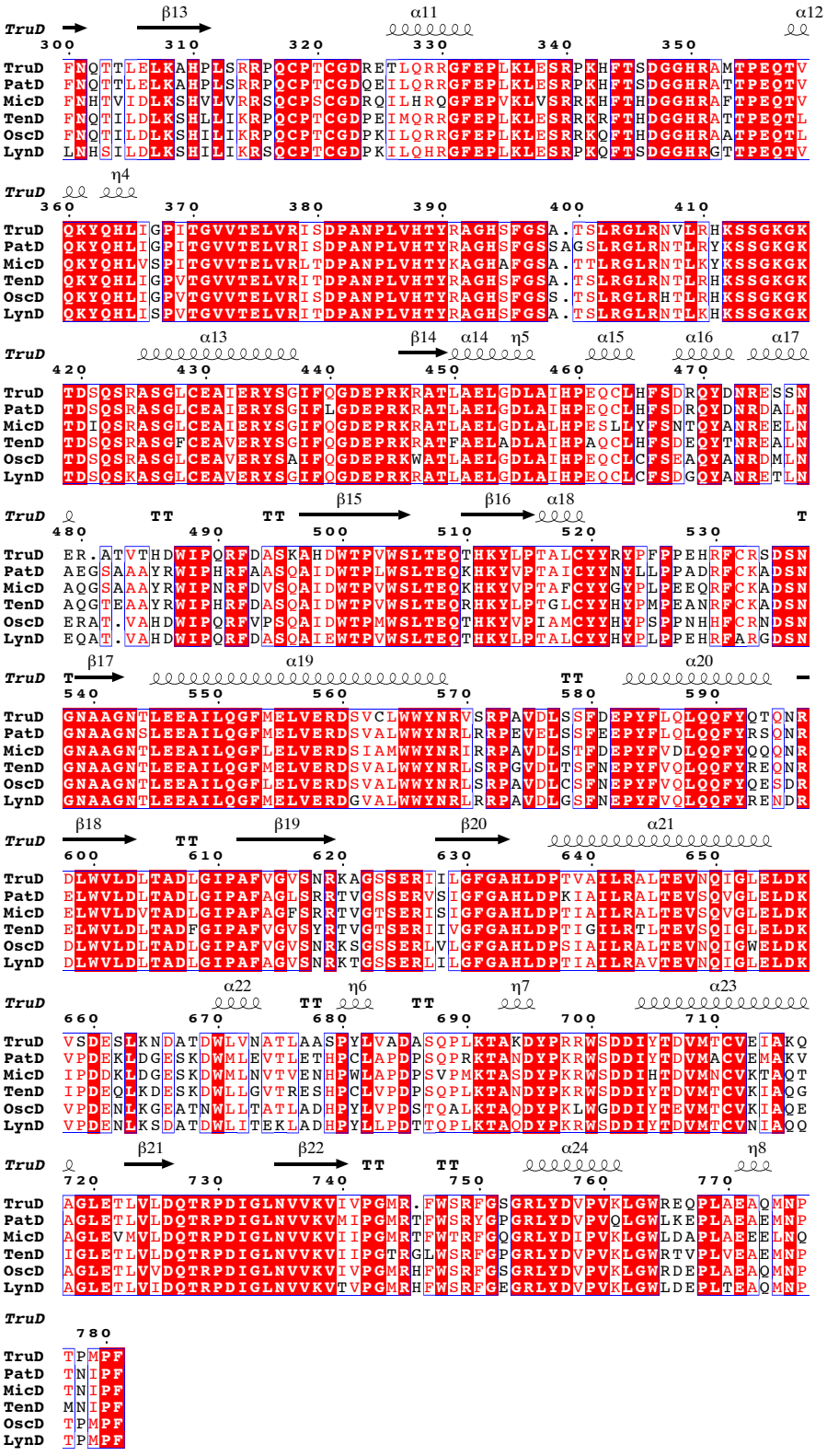

Figure 1.10: Partial sequence alignment of the YcaO domains of heterocyclases TruD, PatD, TenD, MicD, OscD (*Oscillatoria nigro-viridis PCC 7112*) and LynD. Blue boxes show conserved regions; white texts on red backgrounds highlight residues that are identical in all homologues and red texts represent other highly conserved residues. Sequences were aligned with Clustal Omega [84] then plotted and annotated with secondary structures from TruD structure (PDB: 4BS9, chain A) using the ESPript server [85].

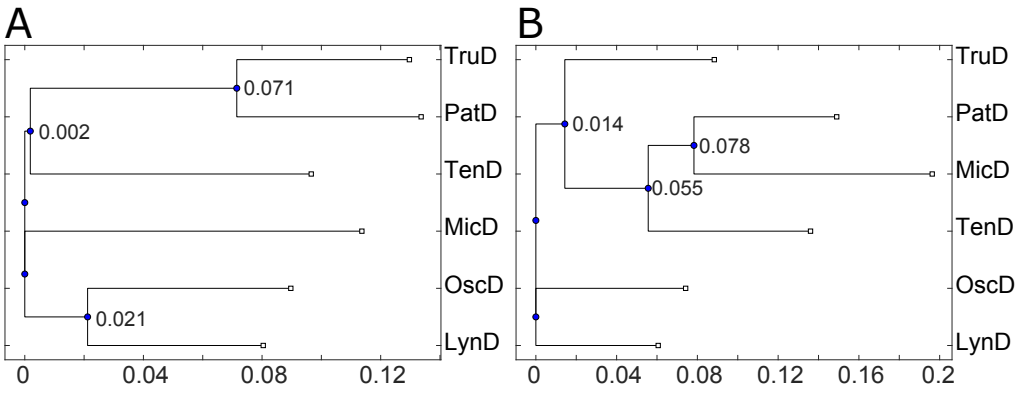

Figure 1.11: Phylogenetic trees of the Cyanobactin Heterocyclases TruD, PatD, TenD, MicD, OscD and LynD. (A) Analysis based on full sequences of the proteins. (B) Analysis based on C-terminal YcaO domains. X axis shows distances from last common ancester. Trees were generated by Clustal Omega and plotted using the Phylogenetic Tree app in Matlab.

which lines the passage that leads the peptide to the nucleotide binding site. A660, T661 and P772 also form part of this entrance. Conformational differences caused by mutation of these residues in the oxazoline-producing homologues may underline their broader substrate scope. Molecule structures of one or more of such enzymes will be required to elucidate the changes.

Directionality is another interesting aspect of the TOMM heterocyclase mechanism. Unlike NRPSs (Figure 1.1A) which operate in a modular fashion, RiPP PTMEs are presented with precursor peptides that may contain multiple modifiable residues. Studies of the heterocyclases BalhD and TruD showed a preferred order of catalysis, where the enzyme starts from the C-terminus and works its way towards the N-terminus [65, 68]. The lantibiotic synthetases LctM and HalM2, on the contrary, exhibit preference for an N-to-C order of reaction (dehydration) [88, 89]. For LctM, it is thought that spatial positioning of the substrate mediated by the leader peptide is combined with decreasing binding affinity for each subsequent intermediate in securing the order of reaction [88]. The reason for these enzymes to maintain an order of reaction might be to prevent the formation of incorrect products during biosynthesis. For example, if an upstream modification enzyme converts the substrate ABC into A*B*C*, and the its downstream enzyme recognises compounds with C* as substrates, an N-to-C order in the first enzyme would preclude the second from encountering the intermediates ABC*, A*BC* or AB*C, which cannot be turned into the correct final product.

1.4 Proteases PatA and PatG

Proteolysis of the N-terminal leader of PatE by heterocyclization, carried out by the protease domain of the enzyme PatA [90], is a pre-requisite for the subsequent macrocyclization reaction. The protease/macrocyclase domain of PatG further cleaves the peptide to remove the C-terminal recognition sequence, as well as cyclising the remaining octapeptide in an head-to-tail fashion [91] .

1.4.1 Structures and Mechanisms of PatA and PatG

The proteins PatA [90] and PatG [91] both harbour a subtilisin-like serine protease domain, with a catalytic triad composed of serine, aspartate and histidine. PatApro cleaves after the 'G(L/V)E(A/P)S' recognition motif (Figure 1.5B), freeing the N-terminus of the core peptide for macrocyclization [90], whereas PatGmac cleaves between the core peptide and the 'AYD' recognition site (Figure 1.5B), and joins the rest of the peptide in a N to C manner [91]. The two enzymes share sequential, structural and and mechanistic similarities (Figure 1.12). The first step in both PatApro and PatGmac reactions are proposed to be the formation of a acyl-enzyme intermediate, as in other serine proteases (Figure 1.12B). In PatA, this will be hydrolysed, resulting in the break of the amide bond. An additional domain in PatG, designated the 'macrocyclization insertion' (Figure 1.12A), is thought to prevent hydrolysis of the first acyl-enzyme intermediate [91]; instead, the deprotonated N-terminus of the core peptide attacks and a second tetrahedral intermediate is formed, which leads to N-C macrocyclization. Unfortunately, the crystal structure for PatGmac lacks electron density for the N-terminal amine of the core peptide sequence [91], which precludes conclusions as to how substrate positioning affects catalysis. It is however hypothesized that the effect is minor, as PatG and its homologues showed notable substrate promiscuity. Peptides containing from 6 to 22 amino acids in the core region and a combination of natural and unnatural amino acids have been successfully macrocyclized, as long as they possess the signature sequence (proline/heterocycle)-AYD [92–94]. In addition, peptide substrates containing up to three triazole groups and sugars were shown to be macrocyclized [95, 96], further expanding the substrate scope of the macrocyclase enzyme to include non-peptidic scaffolds.

1.4.2 Reaction Rates of PatA and PatG

Both enzymes exhibited low turn over rates on various substrates *in vitro*.The hydrolysis reaction catalyzed by PatA is several orders of magnitude slower than observed for other proteases [99] and turnover numbers in the vicinity of 1 day$^{?1}$ for the macrocyclase enzyme have been reported [91, 99, 100]. Tianero and colleagues discovered that PatA

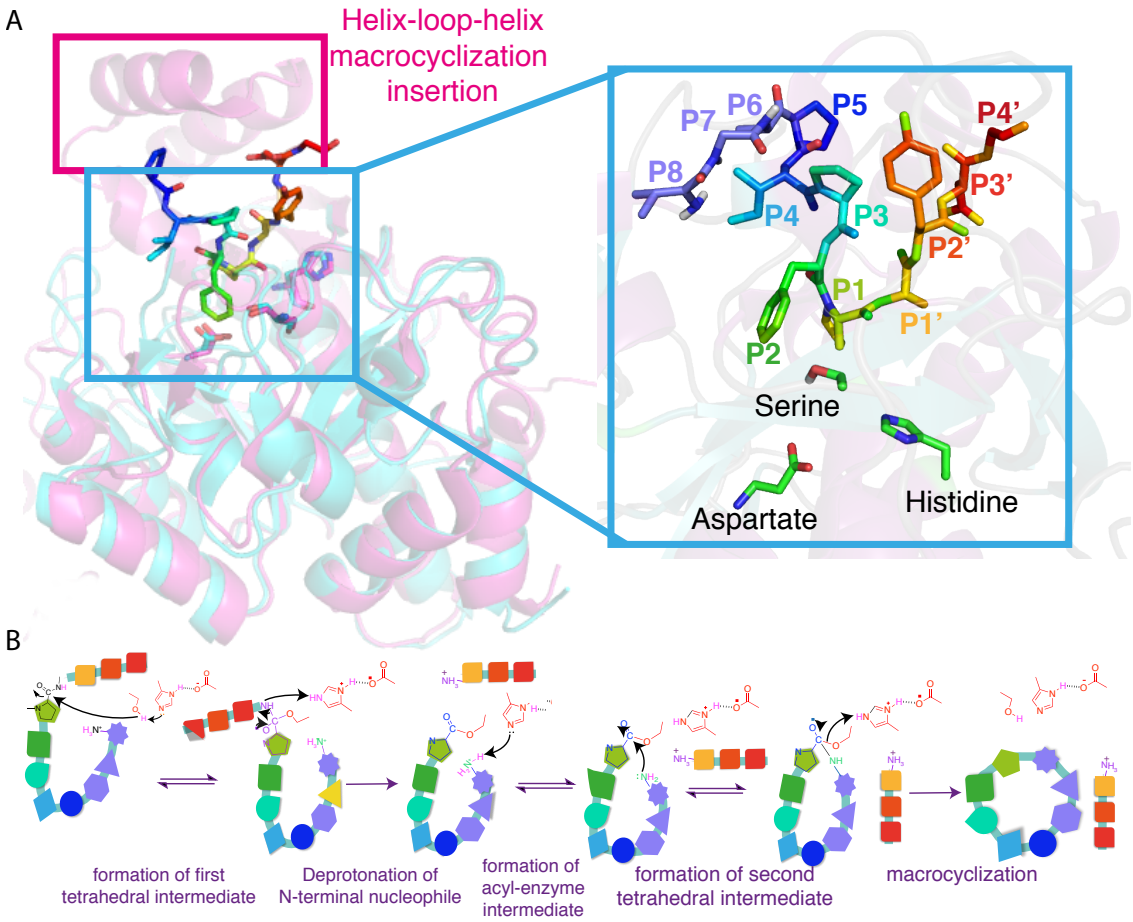

Figure 1.12: Structure and Mechanism of Macrocyclase. (A) The structures of PatG macrocyclase domain (PDB: 4AKT), magenta, in complex with the substrate mimic VPAPIPFPAYDG is overlaid with the apo structure of PatA (PDB: 4H6V), cyan. The pink box hights the helix-loop-helix motif designated the 'macrocyclization insertion' [91]. The blue box shows the active site of PatGmac H618A, in which the site chain for the catalytic histidine (H618) was added; and the three initial residues (VPA) from the core sequence, labeled P6, P7, and P8 were added according to a structure prediction carried out using PEPFOLD [97, 98]. (B) macrocyclization mechanism proposed for PatGmac. The N-terminal amine of the peptide, instead of water, acts as a nucleophile in the formation of the second tetrahedral intermediate. Figure adapted from [45].

was sensitive to reducing agents, commonly used in *in vitro* assays to keep cysteine residues available for heterocyclization [101]. This inhibition could be important in delaying protease cleavage of the leader, which is needed for heterocyclization, and therefore ensuring peptides are heterocyclized before they are cleaved. The authors argued that in cyanobacteria, the native producers of cyanobactins, this process could be regulated by their natural redox cycles [101]. The reason underlying PatGmac's poor catalytic efficiency could lie on the fact that the macrocyclase enzyme seems to trade-off efficiency for a very simple recognition motif. Much faster macrocyclases such as the plant enzymes Butelase 1 (*Clitoria ternatea*), PCY1 (*Saponaria vaccaria*) and the mushroom enzyme GmPOPB (*Saccharomyces cerevisiae*) [102–105] have exhibited higher substrate specificities. Alternatively the macrocyclase enzyme may have evolved to be bottleneck of the process to allow other reactions required to make biologically active compounds to take place [90, 92].

1.5 Azoline Oxidation

Following heterocyclization, thiazolines and oxazolines are in many cases oxidized by a flavin mononucleotide (FMN)-dependent dehydrogenase to yield thiazoles and oxazoles [70]. The oxidase (McbC) was first discovered as part of the thiazole/oxazole synthetase McbBCD involved microcin B17 biosynthesis [70]. In cyanobactin biosynthetic pathways, the oxidase is either a standalone enzyme or a part of the G protein [20]. Like the heterocyclase, the oxidase is absent in pathways that do not require its function, such the trunkamide (*Tru*) pathway, where the final products contains unoxidised thiazolines [106]. Thiazolines arise more frequently than oxazolines among cyanobactins, and also more commonly oxidised [21].

Oxidation of azolines aromatises and flattens the rings, reducing their reactivity [107]. For example, azolines within peptides have been shown to hydrolyse in 10% formic acid, whereas azoles in the same environment stayed intact [68]. Thiazolines and oxazolines are also susceptible to methylation when reacted with methyl sulphate or methyl bromide [108]. Oxidation of thiazolines to thiazoles increase aromaticity more than that of oxazolines to oxazoles [107], and the higher prevalence of thiazoles compared to (methyl)oxazole in cyanobactins [106] could be a reflection of that.

1.5.1 Mechanism of Oxidase Reactions

Oxidation occurs throughout biological systems and serves a variety of functions. Numerous strategies to carry out oxidation reactions enzymatically have evolved in response. Bioluminescence from fireflies is produced when a thiazoline ring in the molecule

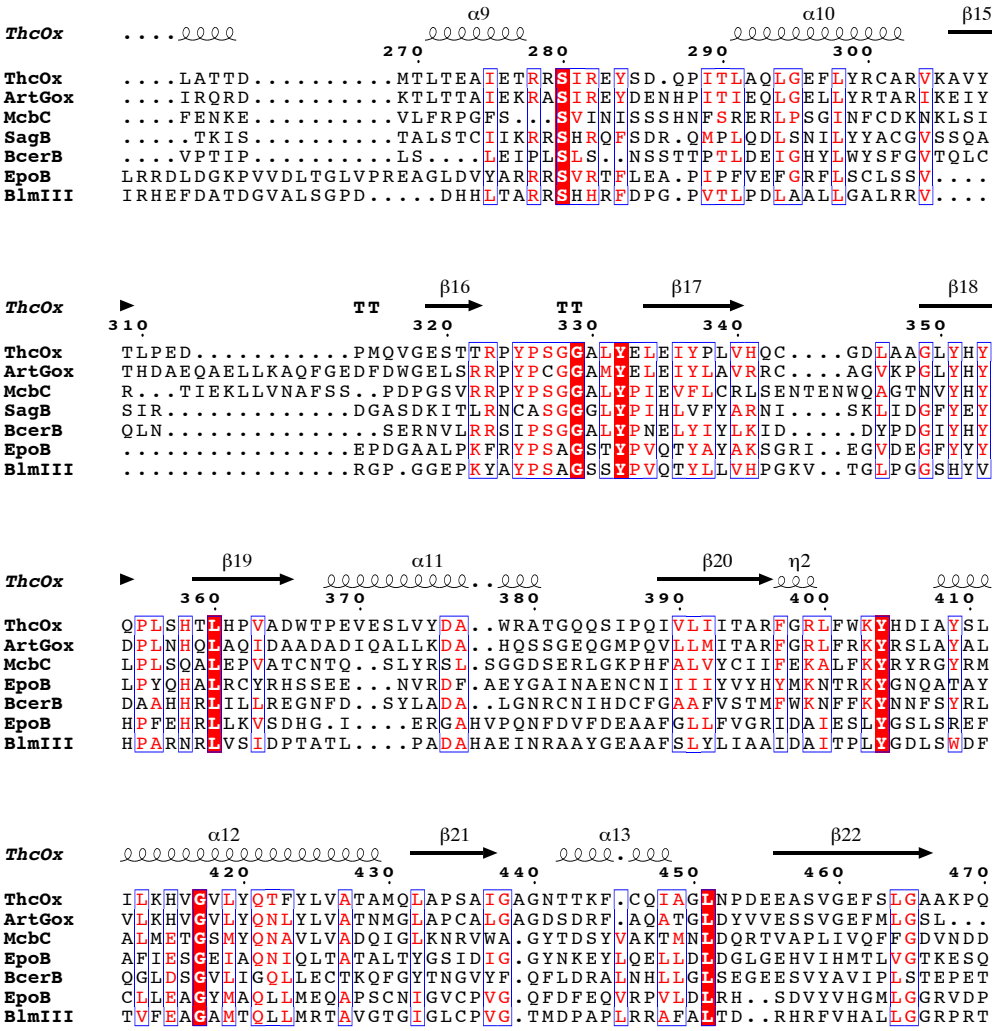

Figure 1.13: Sequence Alignment of RiPP and NRPS oxidases. Cyanobactin oxidases ThcOx (*Cyanothece sp.* PCC 7425) and ArtGox (*Arthrospira platensis*) were aligned with TOMM oxidases McbC (*E. coli*), SagB (*Streptococcus pyogenes*), BcerB (*Bacillus cereus*) and NRPS oxidases EpoB (*Sorangium cellulosum*) and BlmIII (*Streptomyces verticillus*) using Clustal Omega [84]. The alignment file was plotted and annotated with secondary structures from ThcOx structure (PDB: 4LQ5) using the ESPript server [85].

luciferin is oxidised in a ATP/Mg^{2+}-dependent manner, via a dioxetanone intermediate, catalysed by the enzyme luciferase [109]. Rather commonly, iron sulphur clusters or co-factors such as NAD^+, FAD, FMN and quinones are employed in cellular redox processes [110]. Succinate dehydrogenase, a crucial enzyme in aerobic respiration that links the citric acid cycle to the electron transport chain, employ FAD, ubiquinone, as well as several iron sulphur clusters to carry out a series of redox reactions [111]. Azoline oxidases are believed to utilise FMN for redox function. Primary sequence alignments with FMN-bound dehydrogenases from RiPPs and non-ribosomal peptide synthetases show a limited level of sequence homology (Figure 1.13). Most conserved residues are thought to involve in FMN binding [45]. In addition, a lysine-tyrosine motif (KY-motif) is conserved in all the RiPP oxidases (ThcOx, ArtGox, McbB, SagB and BcerB), but is replaced with a 'LY' sequence in the NRP oxidases. Mutagenic analysis of the TOMM oxidases McbC (*E. coli*) and BalhB (*Bacillus* sp. Al Hakam) showed the KY motif to be essential for product oxidation [112]. Analogously, in the related enzyme from *Sulfolobus acidocaldarius* DSM 639, SaciB, mutation of either lysine or tyrosine resulted in complete loss of activity [113]. BcerB from *Bacillus cereus* 172560W showed reduced activity when K185A mutant was tested, whereas both Y186A and the double mutant K185A/Y186A were inactive [113], echoing with the fact that the lysine is not conserved in NRPS oxidases (Figure 1.13). A mechanism for oxidation [114] has been proposed for EpoB (*Sorangium cellulosum*) and BlmIII (*Streptomyces verticillus*), NRPs enzymes catalyzing thiazole formation in epothilone and bleomycin biosynthesis, respectively. In this mechanism (Figure 1.14), the co-factor FMN is reduced twice as the thiazoline is oxidised, becoming $FMNH_2$; molecular oxygen then reoxidises $FMNH_2$, restoring FMN and releasing H_2O_2. The authors observed H_2O_2 using a coupled assay following reactions by EpoB and BlmIII, in agreement with their hypothesis [114]. However, molecular details are not available with regard to catalytic residues and substrate placement.

It is not clear if oxidation occurs before or after cleavage of leader peptide during cyanobactin biosynthesis. Crystals of ThcOx were only obtained when a part of the leader peptide (NILPQQGQPVIR) was added in the co-crystallisation buffer, although that peptide was not seen in the structure [106]. Interestingly, oxidation of heterocycles have been observed on both macrocycles and linear, leaderless peptides, suggesting that the leader peptide is not required for oxidation [87]. On the other hand, the Microcin B17 oxidase (McbC) is part of a heterocyclase/oxidase complex McbBCD [70], which modifies linear precursor with a leader peptide. Indeed, in the case of McbBCD, only the full complex is active, and azolines are not produced when the the substrate was reacted with only the heterocyclase components (McbB and McbD) of it [115]. Other TOMM oxidases such as BcerB, in contrast, are active oxidases on their own, although they still form complex with the heterocyclase when the latter is present [113].

Although cyanobactin oxidases have not been characterised biochemically, studies

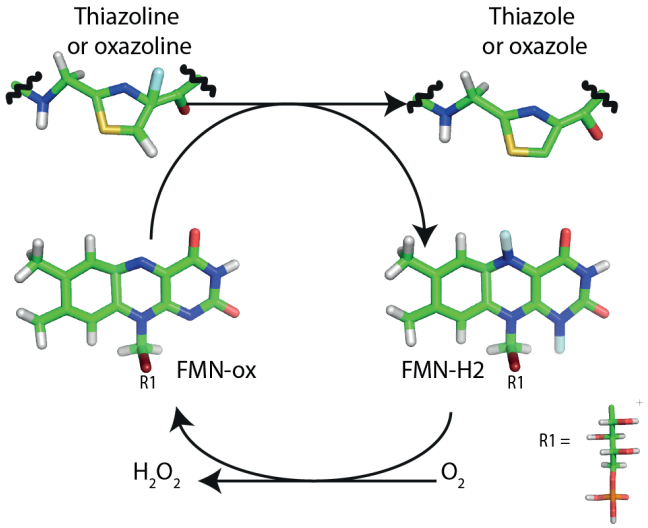

Figure 1.14: Proposed Reaction of Oxidation by EpoB and BlmIII. Figure adapted from [45].

of homologues from other TOMM pathways have yielded insights into the specificity/promiscuity of these enzymes, as well as their order of reaction [68, 87, 113]. Like heterocyclases, oxidases from related organisms were successfully employed interchangeably to produce azoles; however, more distantly related (in sequence terms) enzymes were less efficient, or failed altogether in doing so [68, 71, 113, 116]. Interesting observations were made with regard to the reaction order of oxidase enzyme. When When the oxidase BalhB was present in the same pot with the heterocyclase BalhC/D, azoles were formed on the substrate (BalhA) in a C to N order, similar to that seen for heterocyclization alone [65, 68]; when presented with a heterocyclized intermediate that contains five unoxidised thiazolines, however, BcerB (BalhB homologue) first catalysed the middle thiazoline, and moved onto the N-terminal and C-terminal ones [113].

1.5.2 Structure of the Cyanobactin Oxidase

The X-ray crystallography structure (Figure 1.15) of the standalone enzyme ThcOx (*Cyanothece sp.* PCC 7425) [106] showed a homodimer, each with an FMN-bound oxidase domain and a N-terminal domain (NTD) that contains two 'peptide-clamp' motifs that are prevalent in RiPPs (see section 1.3.4 on page 12). The first peptide-clamp (residues 7-86) is exposed and may involve in substrate recognition, while the second clamp (residues 87-194) is buried at the dimeric interface, bound to the equivalent domain from the other monomer. The NTD is conserved in cyanobactin oxidases, but is absent in other known TOMM oxidases such as BalhB , BcerB and SaciB, which are homologous to the CTD of ThcOx. Mapping conserved residues (identical or similar in all seven enzymes) on the ThcOx structure showed clustering around the FMN binding pocket, as expected;

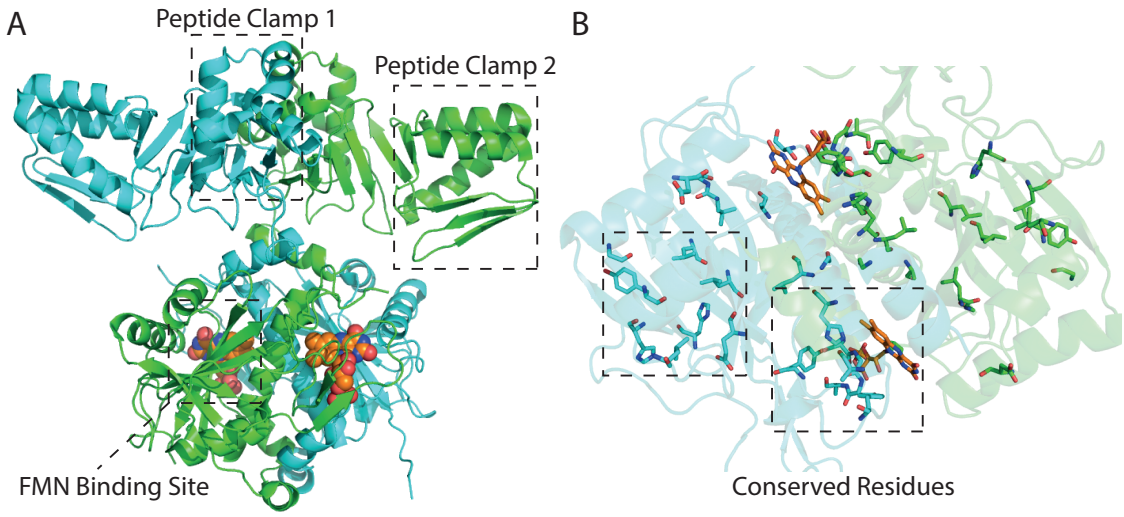

Figure 1.15: Structure of the Cyanobactin Oxidase ThcOx (PDB: 5LQ4). (A) Homodimer of ThcOx, shown in green and cyan, in complex with co-factor FMN, shown in spheres. Each monomer has a N-terminal domain that contains two peptide clamp motifs, and a C-terminal domain homologous to nitroreductases. (B) Residues that are identical or similar across all enzymes aligned in Figure 1.13 are shown in stick representation.

additionally, residues near the exterior of the protein also showed some level of grouping, surrounding strands β18, β19 and helix α10 (Figure 1.15B).

1.6 Prenylase

1.6.1 Prenylation in Cyanobactins

Prenylation contributes to the bioactivity and structural adversity of secondary metabolites, and has been shown to increase cell permeability of peptide substrates [117]. Prenylated cyanobactins (Figure 1.16) feature *O*-prenylation of serine, threonine and tyrosine residues, [16, 21, 50, 118–121], *C*-prenylation of tyrosines [121–123] and tryptophans [124], *O*-geranylation (addition of a C_{10} group) of tyrosines [125], and N-terminal prenylation of short, linear peptides [40, 126]. *O*-prenylation is carried out by a LynF-type protein (Figure 1.17). Specifically, PagF (*Oscillatoria agardhii*) [127] and LynF (*L. aestuarii*) [122] catalyse forward and reverse prenylation on tyrosines, respectively, while TruF1 (*P. dimemni*) [39] reverse prenylates serine and threonine residues, using a dimethylallylpyrophosphate (DMAPP) group as isoprenoid donor. In the LynF pathway (Figure 1.16, aestuaramide), the prenyl group then undergoes non-enzymatic Claisen rearrangement to yield *C*-prenylated products [121–123]. Although PagF-like enzymes had a strong preference for cyclic substrates [101, 122, 127], distantly related linear cyanobactins viridisamides and aeruginosamides undergo prenylation on

the N-terminal amine [40], pointing to the existence of homologues that utilise linear substrates. During aeruginosamides biosynthesis this unusual reaction is carried out by a fused methyltransferase-prenyltrasferase AgeMTPT, which also caps the C-terminus of the precursor with a methyl group [126]. Moreover, PirF, recently reported to forward geranylates a tyrosine residue in piricyclamides using the C_{10} donor geranyl pyrophosphate [128], and KgpF, which catalyses a unique C-3-prenylation of tryptophan in kawaguchipeptins [124] were shown to be active on linear peptides as well as cyclic ones *in vitro*. During cyanobactin biosynthesis *in vivo* the timing of prenylation might be flexible, and could take place alongside macrocyclization.

Interestingly, the prenylase gene is present in almost all cyanobactin biosynthetic pathways, including those that do not have prenylated products, such as the patellamide pathway [20, 40]. Bioinformatics has clustered F enzymes into groups but the relationship between genotypes and phenotypes was supported by limited information [127]. LynF and TruF1 are isolated in their own one-enzyme groups among the 26 enzymes analysed, while the rest formed two major groups - those resembling to PagF and those closer to PatF. PagF-related enzymes, include those with identified prenylated/geranylated products such as AcyF (anacyclamides) [129] and PirF, but also ThcF (cyanothecamides) [130] and ArtF (arthrospiramides) [119], whose known products do not carry prenylation. Even though prenylation in several anacyclamides was confirmed by mass spectrometric analysis, it was not clear on which residues those modifications take place, as some prenylated/geranylated sequences contain threonines/serines, while others have tyrosines or both [129]. PatF is found in the other major group along with enzymes such as TenF (tenuecyclamides) [131] and McaF (microcyclamides) [48], which have also not been reported to prenylate susbtrates. It is unknown why these likely inactive enzymes have been preserved, or have appeared to evolve in the same direction.

1.6.2 Structure and Mechanism of the Cyanobactin Prenylase

Similar to the heterocyclase and the oxidase, the cyanobactin prenylase showed remarkable permissiveness with regard to its substrates [122, 127, 128]. However, LynF's preference for macrocyclic substrates [122] precludes the use of a combination of conserved recognition sequences and hypervariable core as a way of broadening substrate scope, a strategy employed by the heterocyclase [64, 65, 77]. Kinetic analysis of PagF [127], prenylase from the prenylagaramide pathway, showed that that enzymes catalyses tyrosine residues within a peptidic context much more efficiently than L-Tyr on its own, in contrast with canonical small-molecule prenylases. Structures of PagF in complexed with peptide substrates and a DMAPP analogue unveiled the reasons for this selectivity. First, although PagF [127] and PatF [132] both adopt a prenyltransferase fold (Figure 1.18A), comprising a 12-strand beta barrel surrounded by 12 alpha helices, the barrel is open to

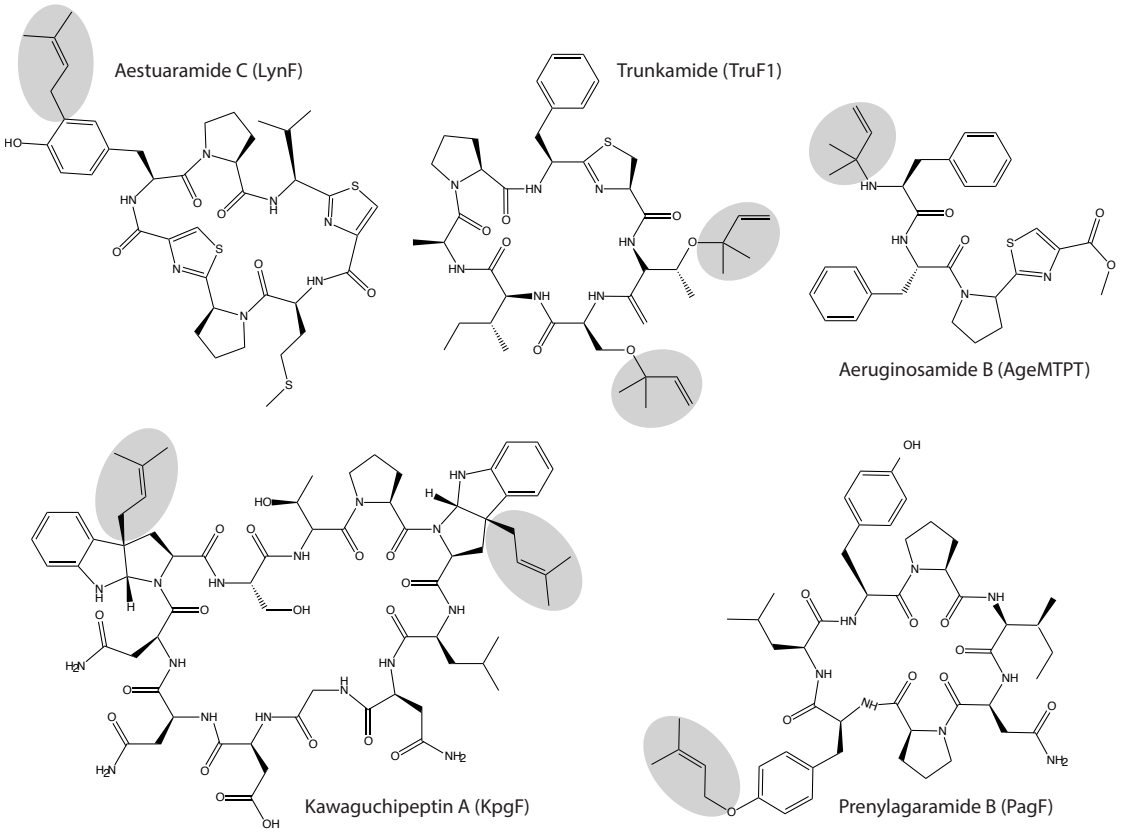

Figure 1.16: Examples of Prenylated Cyanobactins. *O*-prenylation of serine, threonine and tyrosine residues, *C*-prenlylation of tyrosines and tryptophans, and N-terminal prenylation are found in cyanobactins. Prenyl moieties are highlighted in grey.

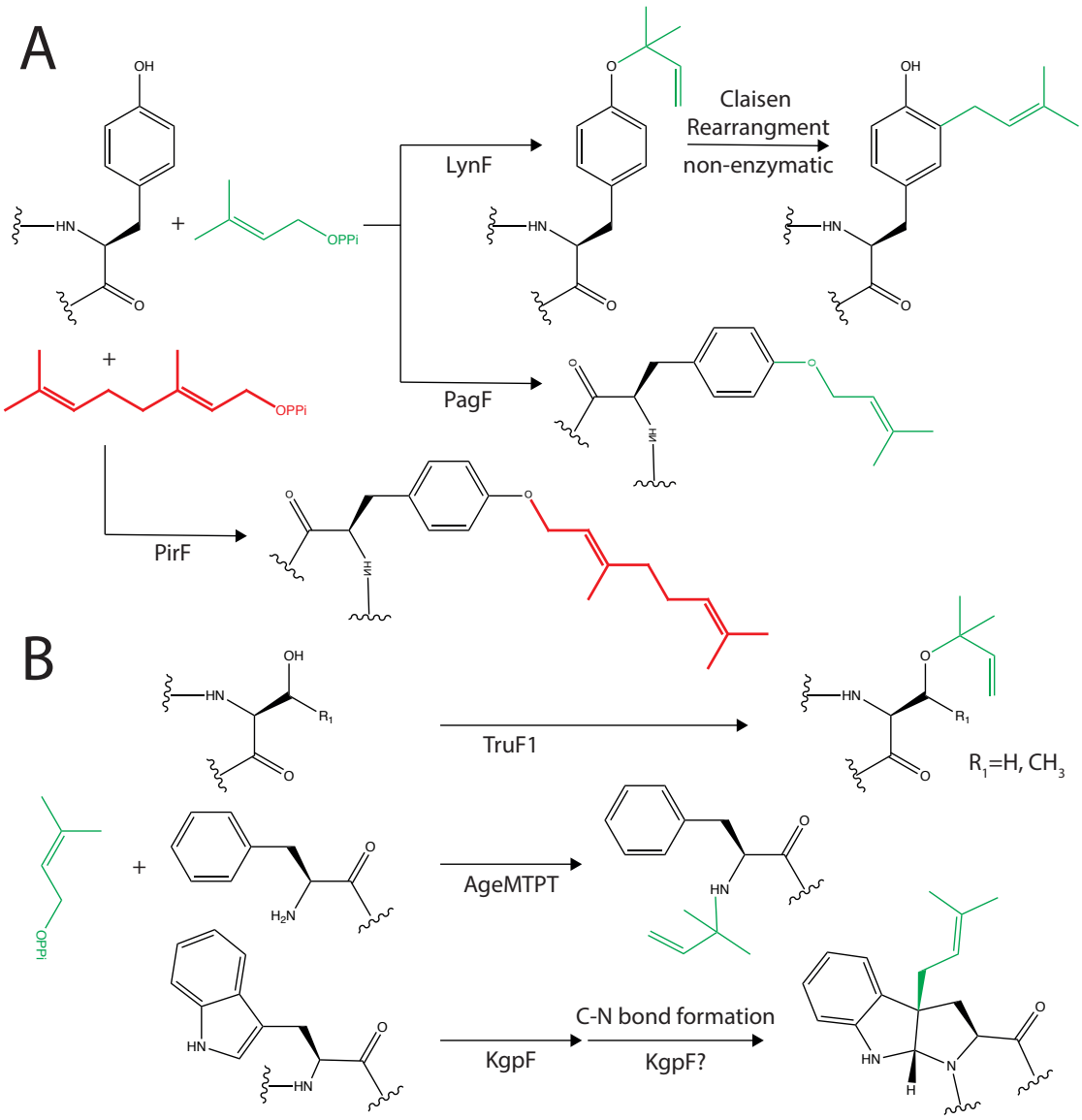

Figure 1.17: (A) *O*- and *C*-prenylation/geranylation of tyrosine. (B) *O*-prenylation of serines/threonines, *N*-prenylation of phenyalanines, and *C*-prenylation of tryptophans.

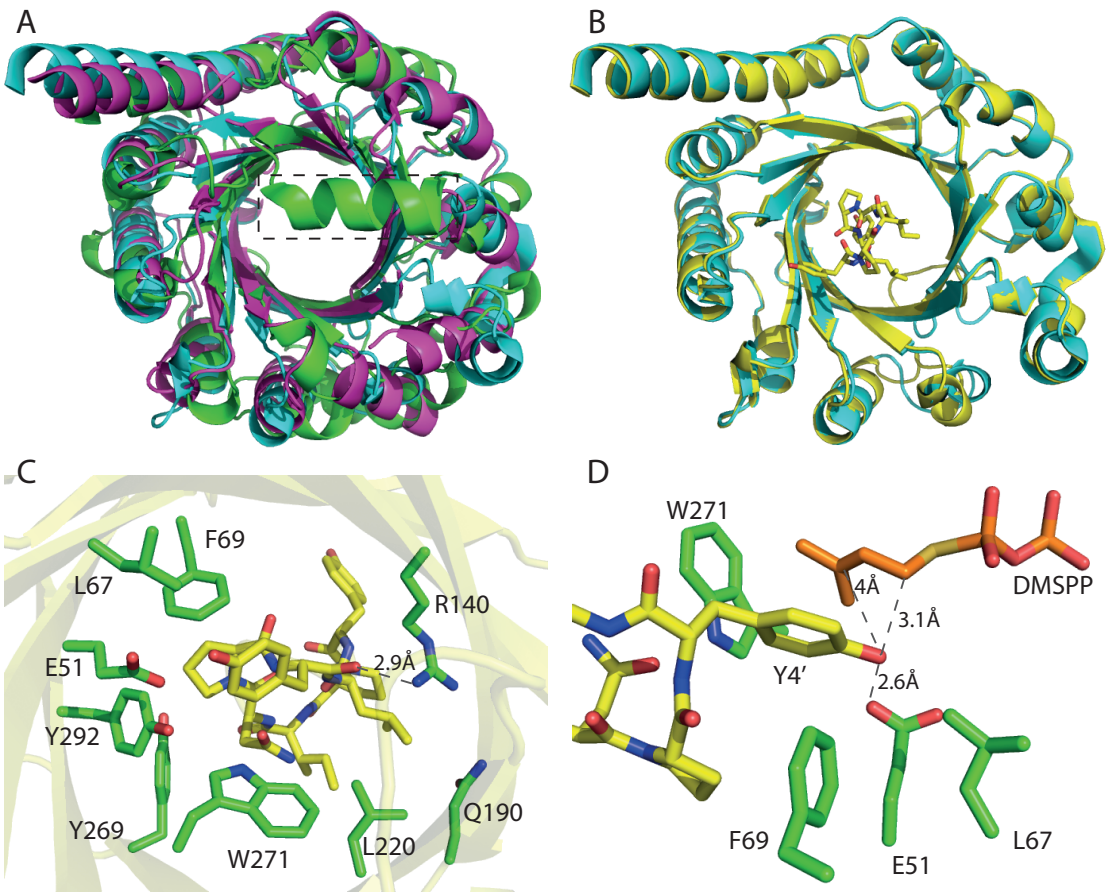

Figure 1.18: Structural basis for the substrate specificity of the Cyanobactin Prenylase. (A) PagF (PDB 5TTY, cyan) is aligned with PatF (PDB 4BG2, magenta) and the bacterial ABBA family member NphB (PDB 1ZCW, green). The capping helix, missing from the cyanobactin prenyltransferases, is highlighted with a black dashed box. (B) The structure of PagF is complex with cyclo-INPYLYP (PDB 5TU6, yellow) shows that the macrocycle forms the bottom of the α/β barrel in the active conformation. (C) The macrocycle, shown as yellow sticks, is enclosed by walls and forms hydrogen bonds or hydrophobic interactions with specific residues from the enzyme, shown in green sticks. (D) Close-up of the PagF active site showing orientation of the substrate analogue DMSPP (orange) relative to the macrocycle (yellow), which favours forward prenylation. Residues from the enzyme are coloured green.

solvents in these enzymes, whereas it is capped by either a helix or a loop of aromatic residues in canonical prenylases [133–135]. In the macrocycle-bound structure of PagF (Figure 1.18B), the peptide plugs the base of the barrel, affording protection to reactive intermediates within the hydrophobic interior [127]. Smaller molecules do not sufficiently shield the cavity, consistent with them being poor substrates. Second, inside the barrel, the macrocycle is secured in place by a range of hydrogen bonds and van der Waals interactions (Figure 1.18C) between cavity wall residues, the modified tyrosine (Y4 of INPYLYP), and amino acids flanking it. Due to structural constraints imposed by the wall, macrocyclic substrates, which bend around the target residue, are preferred to their linear counterparts. Interestingly, despite having two tyrosine residues, only Y4 is prenylated on cyclo-INPYLYP, and complex structures exclusively featured Y4 in the active site, due to energetical favourability of this conformation [127]. Structural restraints also underly the use of tyrosine over serine and threonine, as well as the catalysis of forward rather than reverse prenylation. In cyclo-INPYLYP-bound PagF (Figure 1.18D), Y4 is placed in such a way that the phenol oxygen is 2.6 Å away from E51, which could deprotonate it, and 3.1 Å away from C1 of DMSPP. The distance from this oxygen to C3 of DMSPP is much greater (4.0 Å), consistent with the fact that PagF carries out forward prenylation. Although structures of LynF a d TruF1 have not been obtained, sequence alignment [127] showed mutation of several of the residues shown in Figure 1.18C in these enzymes, which could help them accommodate smaller residues at the active site, or move C3 closer to the prenylated oxygen. To summarise, substrate recognition by the cyanobactin prenylase largely depends on conformational complementarity rather than sequence-specific interactions, and this could have resulted in their promiscuity.

1.7 Unknown Aspects of Cyanobactin Biosynthesis

Despite being one of the best characterised RiPP pathways, patellamide biosynthesis still presents many unanswered questions. The epimerization of the residues next to the thiazoles (the alanine and phenylanine in the case of the peptide ITACITFC) could happen spontaneously under non-physiological chemical conditions [136], but it is not known whether it is catalysed enzymatically in nature. Functions of the nonessential gene products PatB and PatC, the inactive prenylase PatF, and the DUF domains of PatA and PatG are yet to be identified. Variabilities in the number of heterocycles, their oxidation states, and other forms of modifications are only beginning to be explained. Additionally, it remains an open question whether enzymes in the patellamide pathway (for example) form protein-protein interactions (PPIs) which could allow intermediates and substrates to be more efficiently or rapidly funnelled through the pathway. Domains of unknown function (DUF) could act as scaffolds facilitating these PPIs.

Moreover, little is known about the natural destinations of cyanobactins after they are

biosynthesized by cyanobacteria. The natural functions of cyanobactins are unknown, but could be related to symbiosis - profiles of secondary metabolites produced by symbionts are host-specific [137, 138]. Isolation of cyanobactins from ascidians means these compounds are exported to the host organism by the cyanobacterial producer, yet no transportation mechanism has been described. The natural targets of these molecules are also unidentified. It is nevertheless possible to explore their potential medical benefits by screening them against targets of interest. Indeed, efforts have been made to create libraries of cyanobactin derivatives, as described in the next section.

1.8 Current Applications of the Cyanobactin Biosynthetic Machinery

1.8.1 *In-vivo* and *in-vitro* Production of Cyanobactin Derivatives

In-vivo and *in-vitro* procedures have been established to produce cyanobactins and derivatives, creating diversity through varying the genetic composition of the precursor peptide [50, 87, 93, 101, 139, 140]. *In vivo* production by introducing the whole biosynthetic gene cluster into a host organism (usually *E. coli*) is currently ahead of the game in terms of yield and randomised library generation [93, 101], producing up to hundreds of milligrams of cyanobactins per litre of growth medium [101]. On the other hand, *in vitro* methods benefit from increased manipulability of product formation [87, 140], as PTMEs can be individually selected and combined in different ways as needed. The advantages of these biotechnological applications are multifold: employment of the ribosomal machinery allows facile editing of the peptide sequence by means of molecular biology, thereby not only providing a medley of possible product sequences but also enabling the additions of other tailoring facilities, such as peptide cleavage sites [87, 141, 142]; processing of the precursor peptide by PTMEs adds further diversity to the chemical and structural moieties that underly the activities of the final products [16]. Unfortunately, exploitation of the ribosomal route, the very basis for the natural diversity of RiPPs, also poses restrictions on the scope of potential products. A number of non-canonical amino acids, such as selenazolines [143] and p-acetyl-phenylalanine [139] have been introduced into the sequence via medium supplementation and orthogonal tRNA, respectively, but the diversity is limited.

1.8.2 Semi-synthetic Approach Towards Cyanobactin Analogues

A semi-synthetic approach towards cyanobactin derivatives has been adopted by the Naismith group, thanks to the promiscuous nature of the enzymes, especially the

macrocyclases from the cyanobactin family. The macrocyclase PatG only requires its substrate to have a macrocyclization motif AYD (Figure 1.5, page 7), a heterocycle at the position preceding it, and a free N-terminal amine [45]. Chemically synthesised materials that contain azidoalanines, dehydroalanines, triazoles and sugars have been successfully turned into macrocycles using PatG [94–96], making way for further modification by click chemistry, and in the case of sugars, increases membrane permeation by lowering the polarity of the molecule. These non-natural residues are tolerated not only by the macrocyclase, but also by the heterocyclases, allowing the creation of molecules that contain unnatural moieties as well as azol(in)es [94, 95]. The structure-based design of a modified LynD (LynD fusion) was crucial in making this practical [64], as the natural enzyme requires its substrate to have a leader, which makes chemical synthesis difficult, expensive and wasteful. The substrate-bound structure of LynD (Figure 1.8, page 11) showed the interaction between residues QLSSQLAELSEE of PatE and LynD [64]. By measuring the distance between the leader of PatE and the start of LynD in the structure, LynD fusion was created as a recombinant protein containing residues Q21-A37 of PatE fused to the N-terminus of LynD, and is capable of using leaderless substrates that contains only the core sequence (Figure 1.5, page 7) and the macrocyclization motif AYD [64]. As the requirement for leader is bypassed, chemically synthesised peptides (and hydrids) are now readily usable as starting materials. The oxidase and the prenylase may also accept unnatural peptides as substrates, although it has not been directly shown. A summary of the current advances in creating combinatorial libraries of cyanobactin analogues are shown in Figure 1.19. It has to be noted that various methods of synthesis discussed here are primarily laboratory-based, and substantial future work is required to produce them on an industrial scale and test them against therapeutic targets.

1.9 Aims and Objectives

The work constituting this thesis represents continued efforts of exploiting the cyanobactin biosynthetic pathway. Particularly sought after is a tool that installs (methyl)oxazolines into peptide backbones, without the need of a *in-cis* leader sequence. The creation and characterisation of several of such tools (modified enzymes) are detailed in Chapter 2. Beside having a biotechnological interest in the cyanobactin biosynthetic system, we also wish to better understand the fundamental biochemistry of its PTMEs. Chapter 3 and 4 describe, respectively, investigations into the reaction orders of two biosynthetic enzymes of the cyanobactin pathway, the heterocyclase and the oxidase. Chapter 5 then looks into the unusual way via which ATP is consumed by the heterocyclase. Finally, Chapter 6 depicts the characterisation of another biosynthetic enzyme - ATP phosphoribosyltransferase - for which a separate introduction is written.

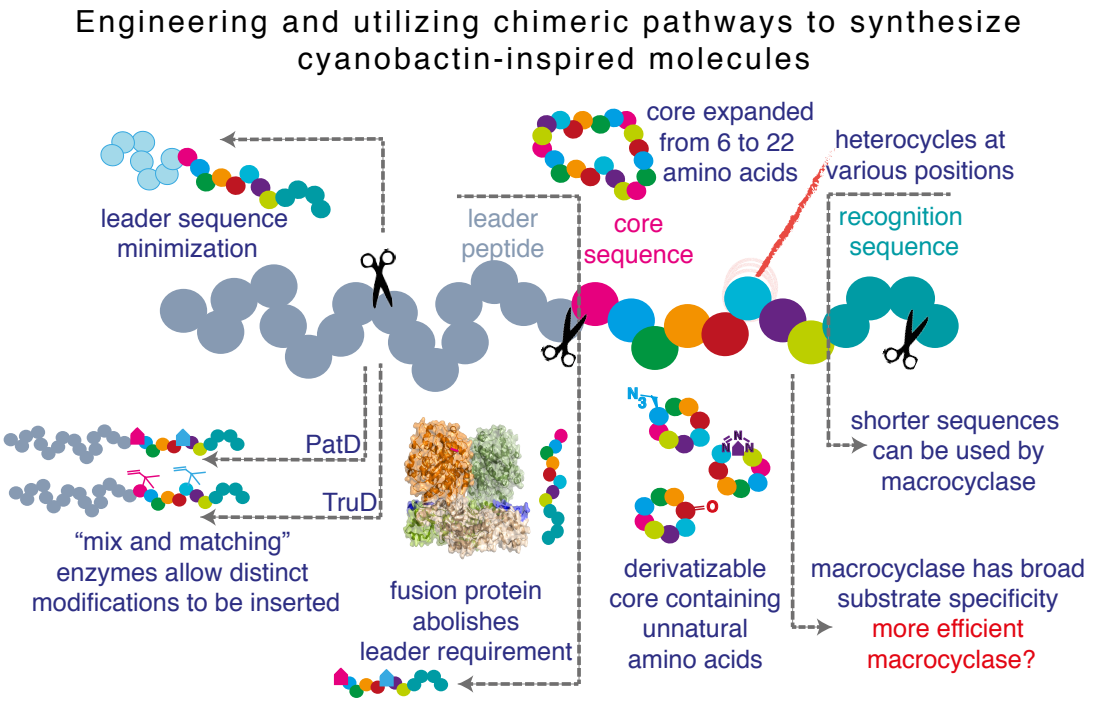

Figure 1.19: Synthesis of Cyanobactin-Like Molecules. Size and sequence of the core peptide, as well as their post-translational modifications can be diversified. In addition, unnatural substrates are accepted by the heterocyclase and macrocyclase. Adapted from [45].

Chapter 2

Engineered Cyanobactin Heterocyclases That Catalyse the Formation of Thiazolines and Oxazolines

2.1 Summary

A range of modified cyanobactin heterocyclases fused with part of the leader peptide of the patellamide precursor PatE were designed, expressed and purified. Several of them were found to capable of catalysing the formation of thiazolines and oxazolines from cysteines and threonines, respectively, when leaderless core peptides containing heterocyclizable residues were supplied as substrates. The reaction condition for catalysing the formation of heterocycles on short peptides was optimised. Variants of the fusion enzymes were compared in terms of their yields and activities.

2.2 Introduction

Leader peptides have a multitude of important roles, such as providing a secretion signal and recognition motifs [78]. The latter is particularly crucial in the biosynthesis of ribosomally synthesised and post-translationally modified peptides (RiPPs), by facilitating the binding between nascent peptides and the their post-translational modification enzymes (PTMEs). Leader peptide of RiPPs may contain several elements that interacts with different PTMEs [16], and could activate the enzymes and/or regulate the positioning

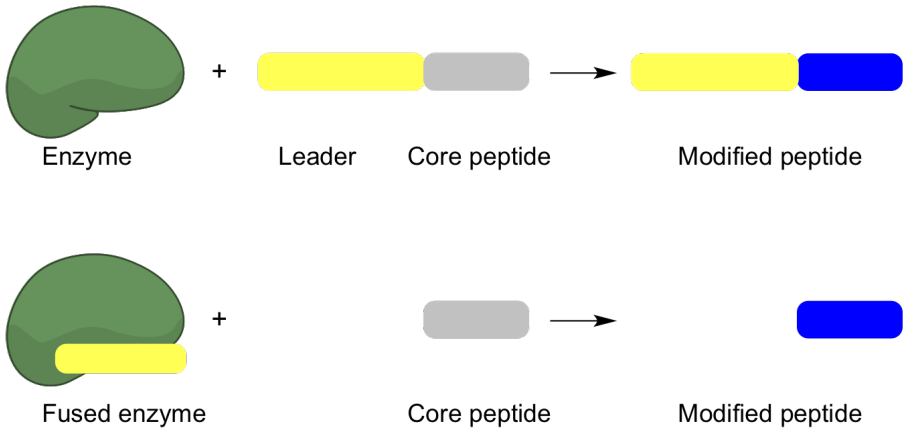

Figure 2.1: A schematic illustration of the fusion enzyme principle.

of the core peptide at the active site [64, 80]. Several studies showed that adding the whole or part of the leader peptide in *trans* could induce the activity of some PTMEs on leaderless peptide sequences [77, 82, 83], and this observation led to the idea of fusing cognate leader peptides to PTMEs, to create constitutively active enzymes. Successful experiments were carried out on the enzymes LctM, HalCE2 (lantibiotic synthetases) [29, 83] and LynD (cyanobactin heterocyclase) [64] (Fig 2.1). In the last case, Koehnke and co-workers attached part of the leader sequence, residues Q21-A37, of the patellamide precursor PatE to the N-terminus of LynD (*Lyngbya sp.*), and produced an enzyme that can heterocyclize multiple cysteine residues on a peptide in the absence of a *in-cis* leader sequences [64]. One major limitation of LynD fusion is that it only modifies cysteines but not threonine/serine residues, which poses restrictions on the diversity of compounds that can be produced. In addition, although the sequence of the fused leader and the length of the linking region was chosen based on structural data, only one construct was made without further optimisation, and there may still be room for improvement. It was unknown whether the other cyanobactin heterocyclases, especially those capable of producing oxazolines, could be modified in the same way to create their constitutively active variants. This chapter describes attempts to extend the fusion strategy to other enzymes. Modifications were carried out on homologues of LynD to create a catalyst that produces oxazolines, and the activities of those engineered enzymes on leaderless peptides are demonstrated.

2.3 Materials and Methods

2.3.1 Preparation of Competent Cells

Competent *Escherichia coli* DH5α (DE3) and BL21 (DE3) cells were prepared based on the Inoue Method [144]. A 10 ml starter culture was made by adding 100 µl of a defrosted

aliquot of existing competent cells into LB medium and incubated at 37 °C, 200 rpm for 12 h. Varied amounts of the starter culture (0.5 ml, 1 ml or 2 ml) were inoculated individually to 200 ml of fresh LB medium and grown for 12-16 h at 18 °C, 200rpm. Optical density (OD) was monitored until one of the cultures reached an OD between 0.5 and 0.7. The culture with the appropriate OD was divided into four 50 ml fractions and placed on ice for 10 min and then centrifuged at 2934 *g* for 10 min at 4 °C. The supernatant was discarded and to each pellet was added 16 ml transformation buffer (10 mM PIPES-KOH pH 6.7, 250 mM KCl, 15 mM $CaCl_2$ and 55 mM $MnCl_2$).The buffer was prepared in advance and stored at -20 °C. Cells were resuspended gently inside Falcon tubes placed on a spinning wheel at 4 °C. After resuspension the cells were centrifuged again at 2934 *g* for 10 min at 4 °C. The pellets were drained and resuspended in 4 ml Inoue buffer each. Resuspended cells were transferred into the same tube to which DMSO (1.2 ml) was added, and the tube was slowly reversed several times. The cells were dispensed into autoclaved and pre-chilled microcentrifuge tubes, flash frozen in liquid nitrogen and stored at -80 °C.

2.3.2 Construction of Expression Plasmids for PatD, MicD, TenD and OscD Fusions

See Fig 2.2 for an illustration of the 'cut-and-paste' strategy used and subsection 2.3.4 for polymerase chain reaction conditions and primer sequences. To start with, a *Hind*III site (AAGCTT) was introduced to the plasmid pJexpress411 carrying the gene encoding LynD fusion [64], by mutating the 6 amino acids directly N-terminal to the LynD coding sequence via site directed mutagenesis [145]. Template DNA was removed by adding 0.5 µl DpnI and incubating at 37 °C for 2 h, and the mixture was purified using a QIAquick Gel Extraction Kit (QIAGEN) following the PCR clean-up protocol. The purified product (10 µl) was used to transform *E. coli* DH5α (DE3) cells by heat shock. Colonies selected from a LB-kanamycin agar plate spread with transformed cells were inoculated individually in 10 ml LB-kanamycin medium, and grown at 37 °C for 16 h. Plasmids were recovered using a QIAprep Spin Miniprep Kit (QIAGEN) and the mutation was confirmed by gene sequencing (GATC). The vector for new fusion enzymes was prepared by sequential digest of LynD fusion carrying the *Hind*III mutation with the restriction enzymes *Hind*III and *Xho*I (Promega), followed by incubation with 5 unit alkaline phosphatase (Promega) for 30 min at 37 °C and purification by 1% agarose gel electrophoresis, to remove the LynD fragment. Agarose gels were run at a constant voltage of 120 V for 45 min in Tris-Acetate-EDTA (TAE) buffer and contained 1 µl of the stain SYBR Safe (Life Technologies) per 35 ml. The DNA was recovered using a QIAquick Gel Extraction Kit (QIAGEN).

PatD (from *Prochloron didemni*) was provided in a pEHISTEV[146] plasmid by the Naismith lab (University of St Andrews). Codon optimised full-length MicD (from *Microcystis aeruginosa*), TenD (from *Nostoc spongiaeforme* var. *tenue*) and OscD (from

Oscillatoria nigro-viridis PCC 7112) with an N-terminal Tobacco Etch Virus (TEV) protease-cleavable His_6-tag were purchased from DNA 2.0 and were supplied in the proprietary pJexpress411 vector. MicD fragment carrying *Hind*III and *Xho*I sites were amplified from the above plasmid by PCR using the primers MicD-f and MicD-r, PatD using the primers PatD-f and PatD-r, and so forth for TenD and OscD. The products were purified by 1% agarose gel electrophoresis, extracted using the QIAprep Spin Miniprep Kit (QIAGEN) and then digested with *Hind*III and *Xho*I (Promega). Ligations with the *Hind*III and *Xho*I digested LynD fusion vector (as described above) were performed at 16 °C for 2 h using DNA ligation kit 2.1 (TAKARA). Plasmids were recovered from *E. coli* DH5α (DE3) cells and the successful cloning of these constructs were confirmed by sequencing (GATC Biotech AG). In the case of TenD and OscD, an optimised variant of the construct that contains leader residues R16-A37 (described in the next subsection) was used as template instead of LynD fusion.

2.3.3 Sequence Optimisation of MicD Fusion

The initial MicD fusion construct has residues Q21-A37 from PatE2K (MDKKN ILPQQ GQPVI RLTAG QLSSQ LAELS EEALG DAGLE ASKIT ACITF CAYDG ELEHH HHHH), five of the amino acid pair GA, followed by the amino acids lysine and leucine which were introduced to facilitate cloning, on the N-terminus of the protein. Extension of the leader sequence to R16, G11 and M1 was achieved by PCR amplification of fragments from PatE DNA borne on a pBMS vector (a gift from Dr Huanting Liu) [146], using the forward primers LeaderR16-f (for leader residues R16-A37), LeaderG11-f (for leader residues G11-A37) or LeaderM1-f (full length leader) and the reverse primer Linker 5GA-r. Shortening or lengthening of the linker to 3GA, 7GA and 9GA was generated by performing PCR with the MicD fusion plasmid as template, using the forward primer FusAmpl-f coupled with reverse primers Linker3GA-r, Linker7GA-r and Linker9GA-r, respectively. All these fragments contained the restriction sites *Nde*I and *Hind*III on their 5' and 3' end, respectively, and were ligated back into the original MicD fusion vector after it has been digested with the same restriction enzymes and gel-purified as described in the last subsection. Lengthening of the linker to 13 GA pairs was achieved by digesting and ligating a gblock (IDT) with 5' *Nde*I and 3' *Hind*III sites to the same vector.

2.3.4 Polymerase Chain Reaction Conditions

Reactions were performed in a Thermocycler (Biorad) using the following programmes. For site-directed mutagenesis, the reaction underwent 12 cycles of denaturation (94 °C for 1 min), annealing (55 °C for 1 min) and extension (68 °C for 1 min 55 s), followed by 3 cycles of 1 min at 95 °C, 1 min at 47 °C and 1 min 55 s at 68 °C. For PCR amplification,

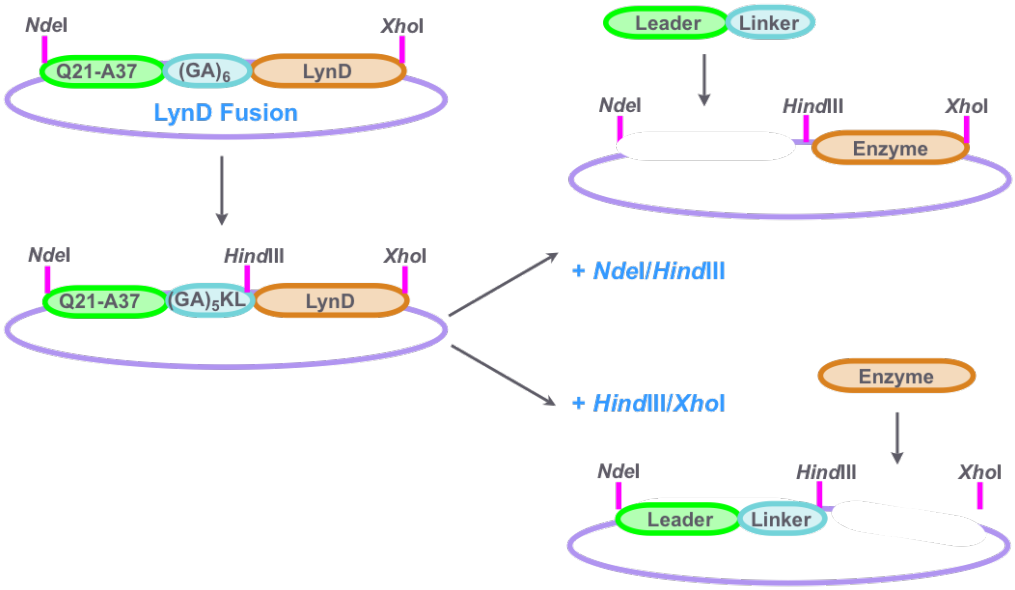

Figure 2.2: 'Cut-and-paste' strategy for construction of varied fusion enzymes.

36 cycles of denaturing at 94 °C for 15 sec, annealing at 60 °C for 30 sec and extending at 68 °C for 15 sec (leader peptides) or 55 sec (full length heterocyclases) were used. The reaction mixtures (50 µl) included 1x KOD polymerase buffer, 0.2 mM of each dNTP, 1.5 mM (site-directed mutagenesis) or 1 mM (other reactions) $MgSO_4$, 1.5 µM of each primer, 10 ng template DNA and 1U KOD hot start polymerase (Novagen). The primers used are listed in Table 2.1.

Table 2.1: List of Primers used in the construction of fusion enzymes

Primers	Sequence (5' - 3')
MutHindIII-f	GGCGCCAAGCTTATGCAATCTACCCCGTTGCTGCAAATT
MutHindIII-r	TGCATAAGCTTGGCGCCGGCGCCTGCACCCGCACC
MicD-f	CTTCTTAAGCTTATGCAGTCGACCCCGCTGCTG
MicD-r	CTTCTTCTCGAGTTAGAACGGGATGTTGGTCTG
PatD-f	CTTCTTAAGCTTATGCAGCCGACCGCGTTGCAGATC
PatD-r	CTTCTTCTCGAGTTAGAATGGGATATTGGTCGG
TenD-f	CTTCTTAAGCTTATGCAGAGCACCACGCTGTTG
TenD-r	CTTCTTCTCGAGTTAGAACGGAATGTTCATCGG
OsciD-f	CTTCTTAAGCTTATGCAGAGCACCACCCTGCTT
OsciD-r	CTTCTTCTCGAGTTAAAACGGCATCGGCGTCGG
LeaderR16-f	CTTCATATGAGCCATCATCACCACCACCACGATTATGACGAAAACCTGTACTTTCAAGGTAGCCGCTTGACTGCTGGTCAGCTC
LeaderG11-f	CTTCATATGAGCCATCATCACCACCACCACGATTATGACGAAAACCTGTACTTTCAAGGTAGCGGCCAACCCGTCATCCGCTTG
LeaderM1-f	CTTCATATGAGCCATCATCACCACCACCACGATTATGACGAAAACCTGTACTTTCAAGGTAGCATGGACAAAAAAAACATTCTA
FusAmpl-f	CTTCATATGAGCCATCATCAC
Linker3GA-r	AAGAAGAAGCTTGGCGCCCGCACCCGCGCCAGCATCGCCGAGAGCTTCCTC
Linker5GA-r	CTTCTTAAGCTTGGCGCCGGCGCCTGCACCCGCACCCGCGCCAGCATCGCCGAGAGCTTCCTC
Linker7GA-r	CTTCTTAAGCTTAGCACCGGCACCGGCGCCGGCGCCTGCACCCGC
Linker9GA-r	CTTCTTAAGCTTGGCGCCTGCGCCAGCACCGGCACCGGCGCCGGCGCCTGCACCCGC

2.3.5 Overexpression and Purification of Heterocyclases

Expression and purification of native MicD were performed based on established protocol.[65] For each protein, 1 µl of plasmid was used to transform 50 µl *E. coli* BL21 (DE3) competent cells. A single colony was picked on a sterile pipette tip, inoculated into 10 ml LB-kanamycin medium, and grown at 37 °C for 16 h. The starter culture (6

ml) was inoculated into 400 ml autoinduction medium prepared according to the Studier method [147] and incubated at 20 °C, 250 rpm for 48 h. Cells were harvest by centrifuging at 8983 *g*, 4°C for 15 min and drained. The pellet was frozen and stored at -80 °C, before being defrosted and resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole and 3 mM 2-Mercaptoethanol (BME)), supplemented with complete EDTA-free protease inhibitor (one tablet per 50 ml of suspension) (Roche) and DNase (0.4 g per gram of cell pellet) (Sigma). The suspension was filtered through a sieve and passaged through a cell disruptor (Constant Systems Ltd), which was equilibrated with lysis buffer and chilled to 4 °C, at 30 kpsi. Cell lysate was cleared by centrifuging at 48384 *g* for 20 min at 4 °C, followed by filtration through a Millipore nitrocellulose membrane with a pore size of 0.45 µm (Sigma). Cleared lysate was applied to a Ni-NTA column (GE Healthcare), equilibrated with lysis buffer, and eluted with elution buffer (lysis buffer substituted with imidazole to the final concentration of 250 mM). The eluate was collected and concentrated where necessary, and subjected to size-exclusion chromatography on a 16/600 Superdex 200 pg column (GE Healthcare) in gel filtration buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)). Where specified, the following steps were included for increased purity. TEV protease (1 mg per 10 mg protein) was added to the eluate from affinity chromatography and the protein-TEV mix was dialysed in a SnakeSkin (Thermo) dialysis tubing with 10 kDa MWCO against 50x volume of low salt buffer (20 mM Tris pH 8.0, 100 mM NaCl and 3 mM BME) for 2 h at 20 °C. The tubing was transferred to fresh buffer and left dialysing at 4 °C for 16 h. After dialysis the protein was applied to a second Ni-NTA column equilibrated in low salt buffer, and the column was washed with 10 column volumes (CV) of low salt buffer. The flow through was passed directly onto a monoQ 10/100 GL column (GE Healthcare) equilibrated in low salt buffer, and the protein was eluted with a linear gradient of low salt buffer and high salt buffer (20 mM Tris pH 8.0, 1 M NaCl and 3 mM BME). The protein eluted at around 250 mM NaCl. The eluate was then subjected to size-exclusion chromatography using buffers and equipments stated above.

Samples of the protein were heated in NuPAGE LDS sample buffer supplement with 50 mM DTT at 100 °C for 5 min and examined by electrophoresis on a Novex NuPAGE SDS-PAGE gel against a unstained Mark12 Protein Standard (Life Technologies) at 220 V for 35 min. Protein bands were visualized by staining with Coomassie Blue and destaining in water. Concentrations were determined using absorbance at 280 nm measured by a Nanodrop (Thermo Scientific) and extinction coefficient calculated by the ExPASy ProtParam tool [148].

2.3.6 Comparison of Expression Levels of MicD Fusion Enzymes

Freshly transformed *E. coli* BL21 (DE3) colonies harbouring pJExp411-MicDfusion plasmids were picked on sterile tips and each inoculated into 10 ml LB-kanamycin medium. The cultures were incubated for 16 h at 37 °C before 6 ml of each was transformed to 400 ml autoinduction medium [147] and grown at 20 °C, 250 rpm for 48 h. 1 ml of each culture was transferred to an microcentrifuge tube and pelleted by centrifugation at 17,000 *g* for 5 min. 1x Bugbuster (Merck Millipore) solution was prepared by mixing 1 ml of 10 x concentrate with 9 ml lysis buffer (see subsection 2.3.5). Each cell pellet was resuspended in 100 µl of 1x Bugbuster solution and incubated at 20 °C, 200 rpm for 20 min. The soluble and insoluble fractions were separated by centrifugation at 17,000 *g* for 10 min, and analysed by SDS-PAGE to obtain expression levels.

2.3.7 Substrates of Heterocyclization Reactions

The PatE substrate used (PatE2K) was expressed as a C-terminally His_6-tagged (non-cleavable), full-length precursor peptide. The plasmid pBMS (a gift from Dr Huanting Liu, University of St Andrews) carrying the *patE* gene was transformed into *E. coli* BL21 (DE3) and recovered from an LB-agar plate containing ampicillin. A single colony was inoculated into 100 ml LB-ampicillin medium and incubated for 16 h at 37 °C, before 10 ml of which was transferred to 1 litre LB medium supplemented with 100 µg per ml ampicillin. The cultures were incubated at 37 °C until their optical density at 600 nm (OD_{600}) reached 0.6, at which point Isopropyl-β-D-thiogalactoside (IPTG) was added to the final concentration of 1 mM. Incubation at 37 °C was continued for 5 h and the cells were harvested by centrifugation at 8983 *g* for 15 min. To obtain the protein, cell pellets were resuspended in urea lysis buffer (8M urea, 20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole and 3 mM BME), at lysed by sonication at wavelength of 15 micrometres (SoniPrep 150, MSE). Cell debris were removed by centrifugation at 48384 *g* for 20 min at 20 °C, and filtration through a Millipore nitrocellulose membrane (Sigma) with a pore size of 5 µm, followed by one with a pore size of 0.8 µm. Clarified lysate was applied to a Ni-NTA column (GE Healthcare), equilibrated with urea lysis buffer, and eluted with urea elution buffer (8M urea, 20 mM Tris pH 8.0, 500 mM NaCl, 250 mM imidazole and 3 mM BME). The eluate was supplemented with 10 mM dithiothreitol (DTT) and incubated at 20 °C for 2 h to reduce all the cysteines, before being subjected to size-exclusion chromatography on a 16/600 Superdex 75 pg column (GE Healthcare) in gel filtration buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP). The protein (about 7 kDa but natively unfolded) was concentrated using a concentrator with 5K MWCO (Sartorius) to 2 mM, flash frozen and stored at -80 °C.

The leaderless peptides were synthesised by Dr B. Nardone (University of St Andrews)

or purchased from commercial sources. The sequences are listed in Table 2.2. Peptides are dissolved in dimethylformamide (DMF) to around 50 mM prior to use.

Table 2.2: List of substrates used in heterocyclization reactions. MW, molecular weight or monoisotopic mass.

Sequence	MW (dalton)	Source
(M1-K42)-ITACITFCAYDGELEHHHHHH	6989	recombinant expression
NH2-ITACITACAYDGE-CONH	1328.6	Dr B. Nardone
NH2-ITACITACAYDGE-COOH	1329.6	Bio-Synthesis, Inc.
NH2-ITACITFCAYDG-COOH	1276.6	Bio-Synthesis, Inc.
NH2-ITACITFCAYD-COOH	1219.5	LifeTein LLC

2.3.8 Heterocyclization Reaction Conditions

The heterocyclization reactions included 100 µM full-length PatE or synthetic peptides and the following components: buffer, sodium chloride, ATP, magnesium chloride and enzyme, each varied during the optimisation process. The details are stated with individual experiments. Temperature and incubation time were also varied in different experiments. Reactions were stopped by precipitating the enzyme with the addition of equal volume of acentonitrile (ACN), followed by vortexing for 20 s and centrifugation at 17,000 g at 4 °C for 5 min [149], or by adding Ethylenediaminetetraacetic acid (EDTA) to the final volume of 50 mM. Samples were analysed by matrix assisted laser desorption ionisation Time-of-Flight (MALDI-TOF) mass spectrometry on a 4800 MALDI TOF/TOF Analyzer (ABSciex), or by liquid chromatography-Electrospray Ionisation mass spectrometry (LC-ESI-MS) on an LCT (Micromass) machine. Data were either visualised using Data Explorer (ABSciex), or exported as peak intensity versus mass over charge (m/z) and plotted using Matlab.

2.3.9 Quantitation of MALDI Data

Relative abundance of species were determined by comparing the areas of the peaks corresponding to each heterocyclization status. Monoisotopic peaks ($[M+H]^+$ and $[M+Na]^+$) were integrated using the 'trapz' function in Matlab, which calculates the peak area under a curve via trapezoidal approximation. For each spectrum 10 values (hydrogen and sodium adducts of peptides with 0-4 heterocycles) were obtained and the each was divided by the sum of all 10. The percentage numbers of H^+ and Na^+ adducts were combined to generate the final abundance for that peptide. Although peptide with different numbers of may ionise differently, horizontal comparison between the same species across samples could be drawn. MALDI-based quantitation has been employed before to monitor heterocyclization by LynD [64].

2.3.10 Overexpression and Purification of *M. tuberculosis* pyrophosphatase

Plasmid pJ411-PPA carrying the gene encoding *Mycobacterium tuberculosis* pyrophosphatase (*Mt*PPase) was kindly gifted by Dr Luiz de Carvalho (Francis Crick Institute). The plasmid was transformed into *E. coli* BL21 (DE3) cells and a single colony was inoculated into 100 ml LB medium supplemented with 50 µg/ml kanamycin at 37 °C for 16 h. The started culture (10 ml) was inoculated into 1 L fresh LB-kanamycin medium and grown at 37 °C, 200 rpm until OD_{600} reached 0.5. The temperature of the incubator was lowered to 20 °C and the cultures were left inside to equilibrate to the desired temperature for 30 min without shaking. IPTG was added to the final concentration of 1 mM to induce the cells, which were left to overexpress the protein at 20 °C for 16 h. Cells were harvested by centrifuging at 8983 *g* for 15 min, drained, and resuspended in PPA Lysis Buffer (20 mM HEPES pH 8.0, 20 mM imidazole, 500 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$), supplemented with complete EDTA-free protease inhibitor (one tablet per 50 ml of suspension) (Roche) and DNase (0.4 g per gram of cell pellet) (Sigma). Cells were lysed and clarified as described in subsection 2.3.5. The lysate was loaded onto a Ni-NTA column (GE Healthcare) equilibrated in PPA Lysis Buffer and eluted with PPA Lysis Buffer substituted with 250 mM imidazole. The eluate was subjected to size-exclusion chromatography on a 16/600 Superdex pg 75 column in 20 mM HEPES pH 8.0. The eluate was concentrated, flash frozen and stored at -80 °C.

2.3.11 Steady-State Kinetics of MicD Fusion Enzymes

Coupled assays were conducted using the EnzChek Pyrophosphate Assay kit (Thermofisher) in 100 µl or 200 µl reactions in 96-well plates and absorbance at 360nm was monitored by a SpectraMax plate reader (Molecular Devices). The assay requires the pH of reactions to be within the range of 6.5 - 8.5. The heterocyclase kinetics buffer contained 100 mM Tris pH 8.0, 50 mM NaCl, 5 mM DTT, 2-amino-6-mercapto-7-methylpurine riboside (MESG), purine nucleoside phosphorylase (PNP) from EnzChek kit, and 10 µM *Mt*PPase. *Mt*PPase was used in place of the PPase included in the assay kit, the quality of which fluctuated. Standard curves for 100 µl or 200 µl reactions in a transparent 96-well plate were established using the pyrophosphate standards provided as part of the kit. A360 values obtained after varied concentrations of PP_i (in duplicates) were incubated for 30 min at 25°C inside the plate reader chamber were plotted against the corresponding PP_i concentrations, and fitted to a linear regression in GraphPad Prism 6. Concentration of P_i was obtained by multiplying the concentration of PP_i by two. The quantities of peptides and enzymes in each reaction varied as individually specified. Reactions were conducted in the heterocyclase kinetics buffer containing 10 mM ATP and 10 mM $MgCl_2$, and were started by the addition of enzyme followed by 10 s of shaking. Readings were

taken every 10 s and after each reading the plate was shaken for 5 s to ensure even mixing of the sample. Initial rate is defined as the rate of reaction when less than 10 percent of the substrate has been consumed. In the absence of peptide, ATP hydrolysis gave a background rate which was subtracted from the observed rates when the enzyme is reacted with peptides. Data was fitted using the Morrison's quadratic equation (Eq 2.1) or its variant (Eq 2.2) [150]. $\Delta[P_i]$, change in inorganic phosphate concentration; Δt change in time; v_{int}, initial velocity; V_{max}, maximum velocity; $[E_T]$, total concentration of enzyme; $[S_T]$, total concentration of substrate; K_m, Michaelis constant; k_{cat}, first order rate constant.

$$\frac{\Delta[P_i]}{\Delta t} = v_{int} = V_{max}\frac{[E_T] + [S_T] + K_m - \sqrt{([E_T] + [S_T] + K_m)^2 - 4[E_T][S_T]}}{2[E_T]} \tag{2.1}$$

$$\frac{v_{int}}{[E_T]} = k_{cat}\frac{[E_T] + [S_T] + K_m - \sqrt{([E_T] + [S_T] + K_m)^2 - 4[E_T][S_T]}}{2[E_T]} \tag{2.2}$$

2.4 Results

2.4.1 Constructing Fusion Enzymes that Produce (Methyl)oxazolines

Due to the inability of LynD fusion to heterocyclize threonines and serines, four proteins were selected for engineering, in an endeavour to create a tool that installs oxazoline heterocycles on leaderless peptides. PatD (*Prochloron didemni*), the heterocyclase involved in the synthesis of patellamides, has been shown to heterocyclize threonines *in vitro* [66, 67, 77]. MicD (*Microcystis aeruginosa*) and TenD (*Nostoc spongiaeforme* var. *tenue*) are YcaO domain-containing proteins found in the gene clusters [21] that lead to production of the (methyl)oxazol(ine)e-containing cyanobactins microcyclamides [48, 49] and tenuecyclamides [131], respectively, and both been previously shown to induce four-fold dehydration on a full-length peptide containing the sequence VTVCVTVC [87]. Finally, OscD (*Oscillatoria nigro-viridis PCC 7112*) is homologue from the biosynthetic pathway of the short, linear cyanobactin Viridisamide A [40]. Although the natural product of OscD only contains one thiazoline, the enzyme has not been tested on substrates that have multiple heterocyclizable residues.

The natural substrates of these enzymes have similar leader sequences (Figure 2.3); in addition, LynD fusion was appended with part of the PatE leader. Therefore it was decided that the same PatE leader was used to engineer all enzymes. PatD, MicD, OscD and TenD were fused with the sequence 'QLSSQLAELSEEALGDA' to their N-

```
OscE      MNKKNILPNPGKPVIRGISGKLPSYLAELSEEALGDAGADAS      42
MicE      MDKKNILPQQGKPVFRTTTGKLPSYLAELSEEALGGNGLEAS      42
LynE      MDKKNILPHQGKPVLRTTNGKLPSHLAELSEEALGGNGVDAS      42
PatE      MDKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEAS      42
TenE      MDKKNILPQQGKPVIRITTGQLPSFLAELSEEALGDAGVGAS      42
          *:******: *:**:*   *:* * **********. *  **
```

Figure 2.3: Leader peptide sequences from PatE, LynD, MicE, TenE and OscE. The first eight residues and the heterocyclase recognition element 'LAELSEEAL' were highly conserved, whereas the rest of the sequence were relatively variable.

termini, connected by the amino acids 'GAGAGAGAGAKL'. The recombinant proteins (His-tagged) were overexpressed in *E. coli* and purified using affinity chromatography and gel filtration chromatography. Purified PatD fusion, MicD fusion, TenD fusion and OscD fusion were reacted with leaderless substrates in the presence of ATP and magnesium chloride, in order to test their efficacies.

Figure 2.4 A shows the size of engineered PatD, MicD and OscD. The proteins ran between the 66.3 kDa and 97.4 kDa molecular weight markers, in agreement with their calculated MW's of between 80 and 95 kDa. Identities of these proteins were confirmed by analysis of trypsin-digested excised gel bands using mass spectrometry. The protein samples shown in these traces were purified by affinity chromatography and size-exclusion chromatography, and were over 90% pure.

Each heterocyclization (cyclodehydration) of a cysteine, threonine or serine residues results in the loss of a water molecule (18 Da), and mass spectrometry has been established as a method of detecting such reactions. Peaks with mass losses of 54 and 72 Da could be observed when 5 µM PatD fusion or MicD fusion was applied to the substrate ITACITFCAYDG (Figure 2.4B and C), corresponding to the heterocyclization of three and four residues, respectively. As the substrate contained two cysteines, it is likely that the threonines in the sequence were also heterocyclized. The reactions were not complete, with the majority of peptides staying in the doubly or triply heterocyclized intermediate state at the time of analysis, but these results nevertheless demonstrated the ability of MicD fusion and PatD fusion to catalyse threonines. OscD, on the other hand, was only able to produce one or two heterocycles after overnight incubation, even at an enzyme concentration of 10 µM (Figure 2.4D).

TenD fusion did not have the expected molecular weight (Figure 2.4A). The plasmid might have been contaminated, or a pre-mature stop codon was formed. The purified protein also showed no heterocyclase activity (Figure 2.4E). Further investigation was not pursued.

These initial results showed that MicD fusion and PatD fusion make promising candidates as biocatalysts for oxazolines. PatD fusion appeared to have a small advantage over MicD fusion (Figure 2.4B and C), but neither would bring the reaction to completion

in the same time scale of the experiment (16 h). During the first purifications, MicD fusion had a yield of 37 mg per litre of autoinduction medium, whereas that of PatD fusion was only 7 mg. Balancing on activity and yield, MicD was selected for further experiments and optimisation.

2.4.2 Optimisation of Reaction Condition for MicD Fusion

The reaction condition developed for heterocyclization by TruD and LynD (10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM ATP, 5 mM $MgCl_2$) did not appear to be optimal for MicD-catalysed turnover of leaderless peptides. Moreover, a decrease in pH was observed (using a pH meter) as heterocyclization progresses, possibly due to the low concentration of the buffering component (10 mM HEPES) in the system. The reaction performed better in alkaline environments - only 14% of peptides were fully heterocyclization (4het) in 100 mM Tris pH 7.0 after 16 h of reaction, but at pH 8.0 the number surged to 41%, and further increased to 53% at pH 9.0 (Fig 2.5A, E). At pH 8.0, raising the concentration of Tris from 20 mM to 200 mM moderately accelerated heterocyclization (Fig 2.5B, E), consolidating the hypothesis that the reaction could suffer from too low a buffering capacity. In accordance, when the concentration of Tris is above 100 mM, the pH remain unchanged incubation for 16 h at 37 °C, whereas a decrease was observed at 20 mM and 50 mM Tris.

Ionic strength also plays a role on regulating the reaction rate, with high salt conditions showing a deleterious effect (Fig 2.5C, E). When an incremental amount of NaCl (50 mM, 250 mM and 500 mM of NaCl) was added to the reaction mixture, the percentage of quadruply heterocyclized peptides reduced from 45.5% to 19% (Fig 2.5E), whereas those of triply and doubly heterocyclized peptides went up from 30% to 37%, and 15% to 30%, respectively.

Finally, the effect of temperature on the reaction was tested. Reaction rates generally increase with temperature; however, as *M. aeruginosa* is a freshwater species [49], its enzymes need to adapt to the changing temperatures of water as the level of daylight varies. Indeed, the MicD reaction appeared to tolerate a wide range of temperatures, with the reaction rate not decreasing much from 37 °C to 25 °C (Fig 2.5D, E). Dropping the temperature to 16 °C moderately slowed down the turnover, while further lowering the temperature down to 4 °C greatly reduced the relative abundance of the 4het species (to 18%) after 16 h incubation, compared to the reactions at higher temperatures (45-56%). It was later noted that the pH of Tris buffers change with temperature and therefore the pH 8.0 buffer made at 25 °C would become pH 7.7 at 37 °C, which may have slowed the rate of the 37 °C experiment down. Nevertheless, reactions at 37 °C suffered more problems such as protein precipitation than those at 25 °C, and 25 °C was therefore set as the standard for further experiments. The pH of Tris buffer is higher at lower temperatures, so the

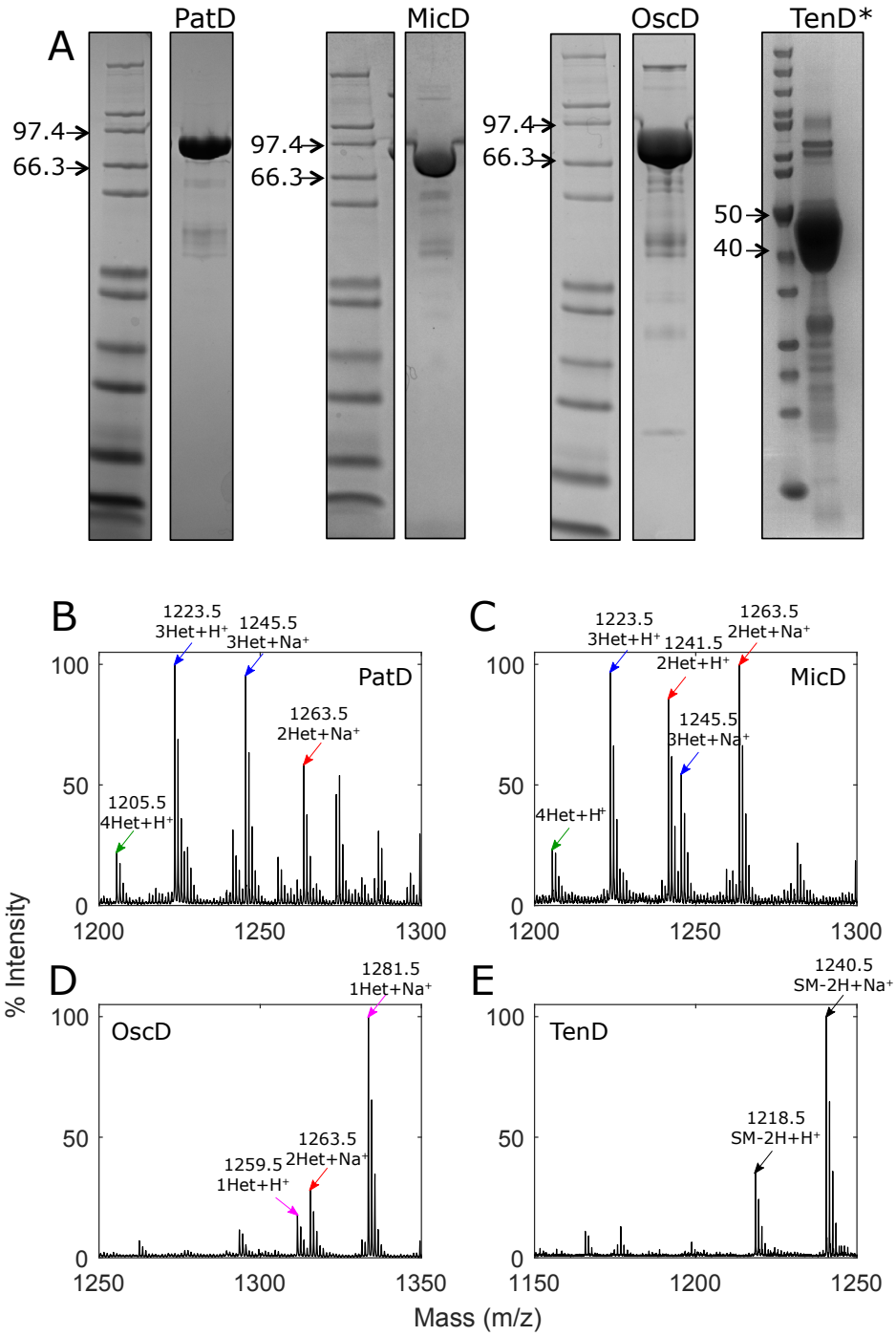

Figure 2.4: Purity and Efficacy of Heterocyclases from Different Homologues. (A)SDS-PAGE traces of purified PatD, MicD, OscD and TenD fusion; (B-E) MALDI-TOF Spectra of leaderless peptides reacted with engineered heterocyclases. (B) the peptide ITACITFCAYDG reacted with 5 µM PatD fusion; (C) the same peptide reacted with 5 µM MicD fusion; (D) The peptide ITACITACAYDGE reacted with 10 µM OscD fusion. (E) The peptide ITACITFCAYD reacted with 10 µM TenD fusion. Reactions were carried out in 10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM ATP, 5 mM $MgCl_2$ and incubated at 27 °C for 16 h. Starting masses are indicated with black arrows, while singly, doubly, triply and quadruply heterocyclized species are indicated with purple, red, blue and green arrows, respectively.

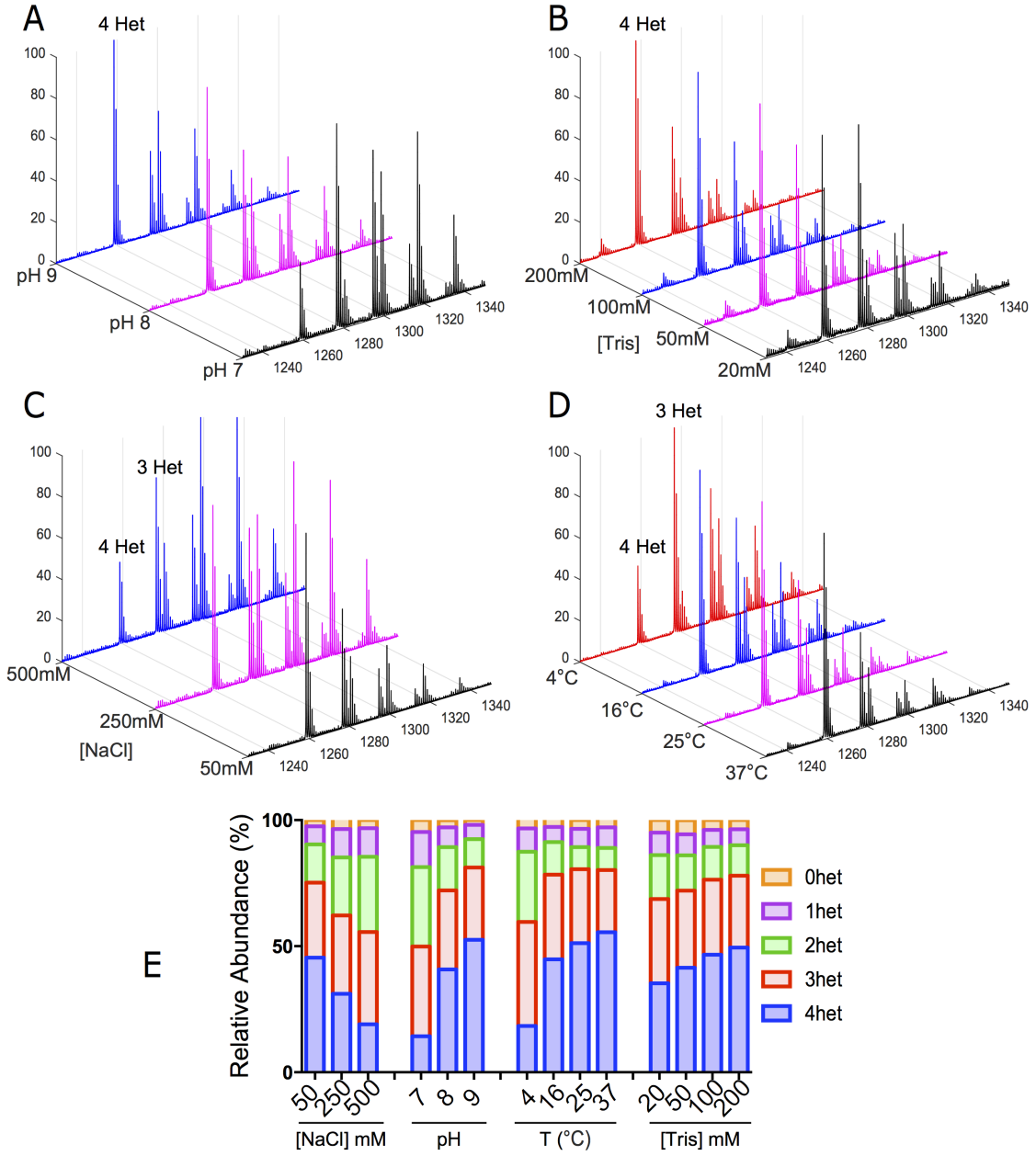

Figure 2.5: Optimisation of reaction condition for MicD fusion, using MicD Q21 9GA. (A) 100 µM peptide (ITACITACAYDGE) was incubated with 20 µM enzyme, in 100 mM Tris pH 7.0, pH 8.0 or pH 9.0, supplemented with 150 mM NaCl,10 mM ATP and 10 mM $MgCl_2$; (B)100 µM peptide (ITACITACAYDGE, C-amide)was incubated with 20 µM enzyme, in 20, 50, 100 and 200 mM Tris pH 8.0, 150 mM NaCl, 10 mM ATP and 10 mM $MgCl_2$; (C) 100 µM peptide (ITACITACAYDGE) and 20 µM enzyme reacted in 100 mM Tris pH 8.0, supplemented with 50,150, 250 and 500 mM NaCl,10 mM ATP and 10 mM $MgCl_2$. (D) 100 µM peptide (ITACITACAYDGE) was incubated with 20 µM enzyme, in 100 mM Tris pH 8.0 supplemented with 50 mM NaCl,10 mM ATP and 10 mM $MgCl_2$ at 37, 25, 16 and 4 °C for 16 h. Peak intensities are normalised and plotted on Z axes against mass in m/z on X axes in Matlab. Data from related experiments are stacked with each other along the Y axes. (E) Relative abundance of each heterocyclization states calculated from (A)-(D). This is a replot of relative peak sizes and not a rigorous quantitation.

reduction in reaction rates at 16 °C and 4 °C should not have resulted from temperature-associated pH change, based on the results shown in Fig 2.5A).

It should be noted that peptides with different number of heterocycles may ionise with varied strengths, thus the numbers in Fig 2.5E do not represent the true abundance of each species in a sample. Rather, the distribution of species was compared horizontally between different samples to determine which conditions produce more heterocycles after the same duration of reaction. It is possible that peptides with higher numbers of heterocycles were underrepresented in the MALDI spectra. Figure 2.6 contrasts the spectra of the same sample run on LCT (ionisation by electrospray) and MALDI - even though the quadruply heterocyclized (4het) species generated the highest ion counts in both spectra, the former gave a 4het:3het ratio of about 5:1 in terms of height, and showed residual amount of 2het species, whereas the latter had a much lower 4het:3het ratio at 2:1, and more prominent 2het peaks, similar to the spectra shown in Figure 2.5. ESI-MS is widely thought as a more suitable method for the quantitation of small molecules than MALDI [151], and the LCT spectrum (80% quadruply heterocyclized) is thus likely a better reflection of the true distribution of peptides in the reaction sample.

A problem that frequently occurred during the initial stage of the optimisation process was that the peptide was not heterocyclized, despite the reaction mixture having all components (buffer, salt, ATP, magnesium, enzyme and substrate). This issue almost aways concurred with an ion species that has a mass 2 Da less than the starting material being present instead of the expected $M+H^+$ species in mass spectra It was therefore possible that an internal disulphide bridge has formed between the two cysteine residues (as of the sequence ITACITFC), which precludes the cysteines from being heterocyclized. Indeed the problem was solved after including a reducing agent such as TCEP or DTT in the reactions (5 mM DTT determined to be optimal). The reason why the reaction still sometimes proceeded in the initial tests, without additional reducing agent, may be due to the fact that a small amount of TCEP was supplemented in the gel filtration buffer of the enzymes, which degraded over time.

2.4.3 Optimising the Sequence of MicD Fusion

In addition to the chemical conditions of the reaction, it is possible that the sequence of the MicD fusion itself could be optimised to further increase the efficiency of heterocyclization of leaderless peptides. Moreover, it would also be desirable to improve the yield and stability of the protein. Multiple MicD fusion constructs were made and those which have been grown and overexpressed in *E. coli* are listed in Table 2.3, which states the named designated to each construct, a description of its modification, its average yield where applicable, and its molecular weight in kDa. Six fusion constructs were purified, MicD Q21 5GA, MicD Q21 7GA, MicD Q21 9GA, MicD R16 5GA, MicD G11 5GA and MicD

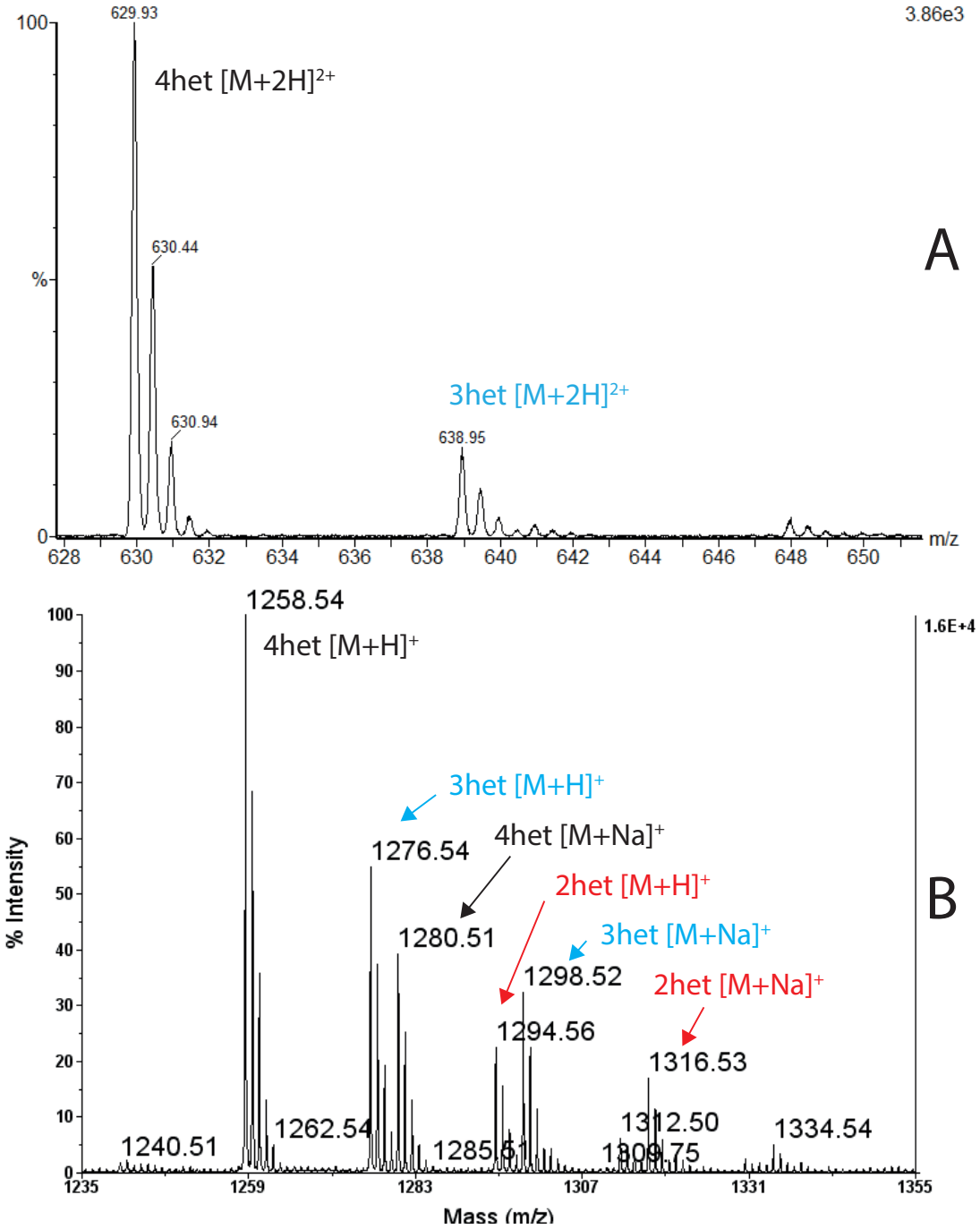

Figure 2.6: Discrepancy in MALDI and LCT Spectra Results. (A) LCT spectrum and (B) MALDI spectrum of ITACITACAYDGE reacted with MicD Q21 9GA in 100 mM Tris pH 8.0 supplemented with 150 mM NaCl, 10 mM ATP, 10 mM $MgCl_2$ and 5 mM DTT, for 16 h at 37 °C.

Table 2.3: Nomenclature, yield and parameters of MicD fusion constructs compared to the native enzyme. Yields are as of protein per litre of autoinduction medium. MW, molecular weight in kDa including an N-terminal His_6-tag and TEV protease site, calculated by ExPASy ProtParam tool.

Construct	Fused sequence	Linker	Average Yield	MW
MicD native	N/A	N/A	15 mg	88.8
MicD M1 5GA	M1-A37	$(GA)_5KL$	1 mg	93.6
MicD G11 5GA	G11-A37	$(GA)_5KL$	30 mg	92.4
MicD R16 5GA	R16-A37	$(GA)_5KL$	55 mg	91.9
MicD Q21 3GA	Q21-A37	$(GA)_3KL$	N/A	91.1
MicD Q21 5GA	Q21-A37	$(GA)_5KL$	38 mg	91.4
MicD Q21 7GA	Q21-A37	$(GA)_7KL$	15 mg	91.7
MicD Q21 9GA	Q21-A37	$(GA)_9KL$	42 mg	92.0
MicD Q21 13GA	Q21-A37	$(GA)_{13}KL$	N/A	92.3

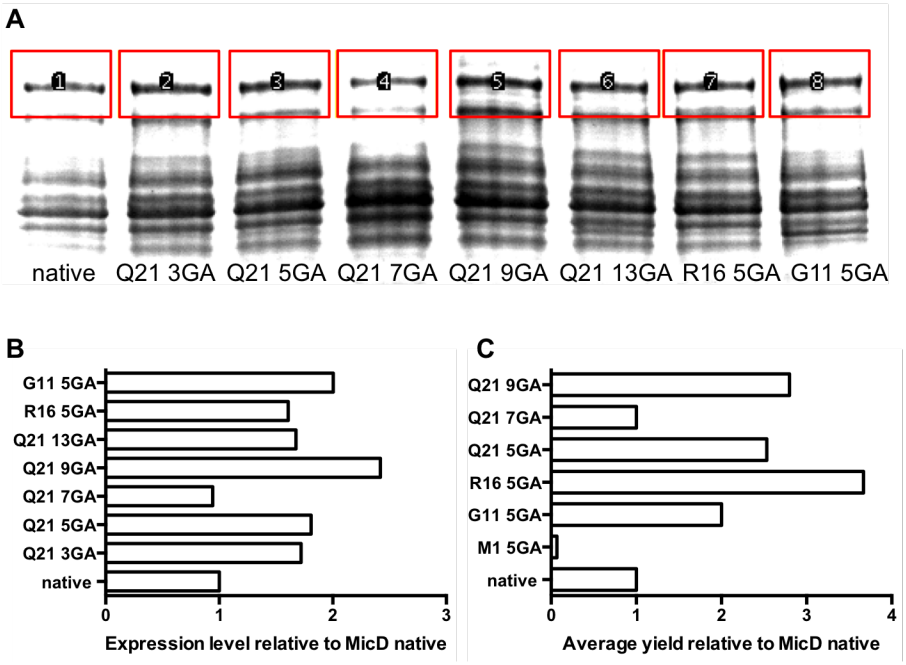

Figure 2.7: Expression Levels and Yields of MicD and Fusion Variants. (A) SDS-PAGE showing soluble expression levels of different MicD variants. Red boxes indicate areas selected for integration in Image J. The image is part of a gel which is shown in Appendix B (Figure S2), and the contrast has been adjusted to aid band selection and quantitation. (B) Relative expression levels as integrated from (A). (C) Normalisation and replot of yields recorded in Table 2.3.

M1 5GA, along with the native enzyme. Native MicD protein had a moderate yield of 15 mg per litre of autoinduction medium. In contrast, average yields of the fusion proteins, recorded during protein purification, ranged from 38 to 55 mg per litre, with the exception of MicD M1 5GA (full leader sequence attached), which expressed poorly at 1 mg per litre. In accordance with the recorded yields (Table 2.3, Figure 2.7C), fusion proteins MicD Q21 9GA, MicD Q21 5GA, MicD R16 5GA and MicD G11 5GA also had higher expression levels in a small scale test performed in autoinduction medium (Figure 2.7A and B).

Figure 2.8 compares the heterocyclase activities of native MicD and its fusion variants. MicD native is capable of fully heterocyclizing the core sequence ITACITFC, provided that it is within a full-length peptide that has a leader sequence (Figure 2.8A). However, without the leader, MicD native only installs 1 heterocycle on the same core sequence (Figure 2.8B). Interestingly, it is capable of dehydrating the sequence ITACITAC twice (Figure 2.8C), suggesting the differences in sequence affects the activity of the enzyme. The native enzyme was nevertheless unable to fully heterocyclize the substrate over the course of the experiment. The fusion enzyme MicD Q21 5GA (Figure 2.4C), in contrast, installs four azolines on the same peptide. In addition, MicD Q21 7GA, MicD Q21 9GA, MicD R16 5GA and MicD G11 5GA (Figure 2.8E-H) all demonstrated the ability to fuly heterocyclize leaderless peptides. MicD M1 5GA, the full-leader fusion, was inactive (Figure 2.8D) as well as having a low yield (Table 2.3). As discussed in section 2.4.2, signal intensities in MALDI do not accurately represent peptide concentration. To find out which construct was the most active, a more quantitative method was sought after, and kinetic assays were employed as such.

Quantitative comparison of the activities of the enzymes MicD Q21 5GA, MicD Q21 7GA, MicD Q21 9GA, MicD R16 5GA, MicD G11 5GA were achieved through steady-state enzyme kinetics using the EnzChek pyrophosphate assay. The assay employs purine nucleoside phosphorylase (PNP), which uses MESG (2-amino-6-mercapto-7-methylpurine riboside) and inorganic phosphate (P_i) to make ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine, a compound that has an maximum absorbance at 360 nm. One molecule of MESG reacts with one molecule of P_i, and as a result, changes in A360 can be used to quantify the amount of phosphate present. As previous studies showed that heterocyclases utilise ATP and release either ADP and P_i or AMP and PP_i, a preliminary reaction was set up, in the presence or absence of an excess amount of pyrophosphatase (PPase), to see whether MicD makes phosphate or pyrophosphate. PPase from *M. tuberculosis* (*Mt*PPase), overexpressed in *E. coli* and purified as a His-tagged protein (Figure 2.9A), was used in place of the PPase included in the assay kit, the quality of which fluctuated. Phosphate is released, as shown by an increase in A360, in the absence of a peptide substrate (Figure 2.9B, green line). Adding PPase (purple line) had minimal effect on the readings, an indication that PP_i was not made in large numbers by the background ATP hydrolysis. Adding the peptide resulted in an increase

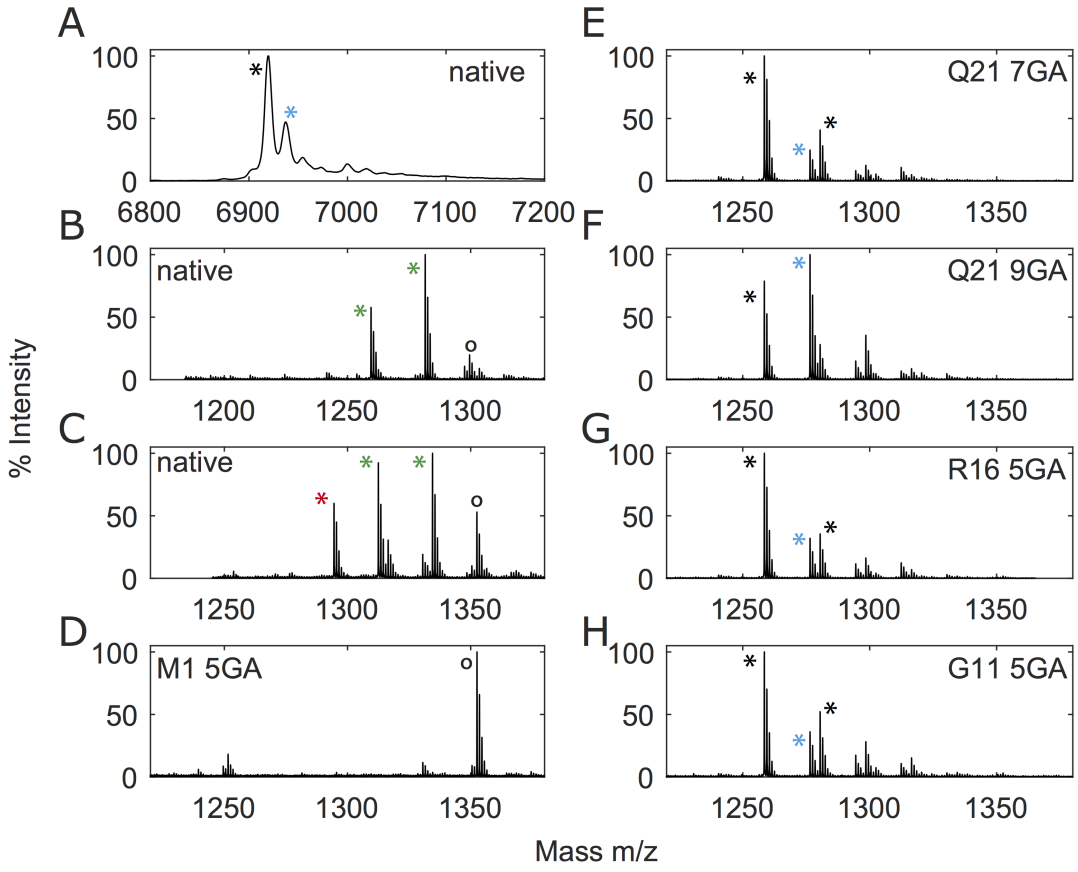

Figure 2.8: Activity of MicD native and MicD fusions. MicD native was reacted with (A) full-length PatE2K, (B) ITACITFCAYDG and (C) ITACITACAYDGE; (D-H) Fusion enzymes MicD M1 5GA, MicD Q21 7GA, MicD Q21 9GA, MicD R16 5GA and MicD G11 5GA were reacted with the peptide ITACITACAYDGE. Reactions containing 5 µM of enzyme and 100 µM of substrate were carried out in 100 mM Tris pH 8.0 supplemented with 50 mM NaCl,10 mM ATP and 10 mM $MgCl_2$. Black circles annotates the starting material in each spectrum. Green, red, cyan and black asterisks annotates singly, doubly, triply and quadruply heterocyclized peptides, respectively.

in P_i production (blue line), as the slope of the linear regression curve more than doubled. PP_i was also produced, demonstrated by a further increase in the rate of the coupled assay when PPase is included, with the slope of the red line (+peptide, +PPase) being over six times steeper than that of the green line (no peptide, no PPase). Two conclusions can be made from the experiment shown in Figure 2.9B. First, heterocyclization catalysed by MicD releases both P_i and PP_i as products; in subsequent reactions, PPase was included so the total amount of phosphates was monitored, including P_i and PP_i. Second, MicD hydrolyses ATP in the absence of peptides. As a result, a peptide-free, background rate was measured for each enzyme, and subtracted from the readings of peptide-containing reactions. The phosphate chemistry of MicD is discussed in more details in Chapter 5.

Calibration curves (Figure 2.9C) were established for 100 µl and 200 µl reactions using the pyrophosphate standards provided with the kit. The pH of the buffer was set at pH 8.0 as it was suitable for both the heterocyclase reaction and the coupled reactions. Linear regression of the 100 µl sample gave a slope of 0.003149 ± 0.0001244 whereas that of the 200 µl samples gave a slope of 0.007477 ± 0.0002630. During coupled assays, changes in A360 over time were monitored, and the rates were divided by the slopes stated above to give changes in PP_i concentrations over time, and then timed by two to give rates in $\Delta[P_i]$/time.

The peptide ITACITFCAYD (10 µM to 500 µM) was reacted with five different fusion enzymes (5 µM each). The initial rates were plotted against substrate concentrations and fitted to Morrison's Quadratic Equation (Equation 2.1). Data points and regression curves are shown in Figure 2.9D. V_{max} and K_m values were calculated using the equation, while k_{cat} was derived from V_{max}. The kinetic parameters are listed in Table 2.4. All fusion enzymes exhibited similar kinetics, with k_{cat} values in the range of 0.0062 s^{-1} to 0.0087 s^{-1} and K_m values ranging from 45 µM to 98 µM. As 100 µM was the concentration of peptide routinely used in catalysis, k_{cat}/K_m was calculated as an indicator of the catalytic efficiency. The first fusion enzyme, MicD Q21 5GA, was the slowest of the enzymes tested, while other fusion enzymes showed a two- to three-fold increase in k_{cat}/K_m. However, all kinetic parameters were within the same orders of magnitude. On the other hand, when the native native enzyme was reacted with leaderless peptides, the rate became too small to measure. These results confirmed that fusion enzymes had increased affinities for leaderless peptides and could perform tasks that were proven difficult for the native enzyme.

Determination of the kinetics of PatD was also attempted; however, the background hydrolysis rate (Figure 2.9E, blue empty circles) was much higher compared to that of MicD (red empty squares). With only a moderate increase of rate when peptide is present (blue filled circles), it was difficult to distinguish errors from actual readings, especially at low concentrations. As a result, kinetic parameters were not obtained.

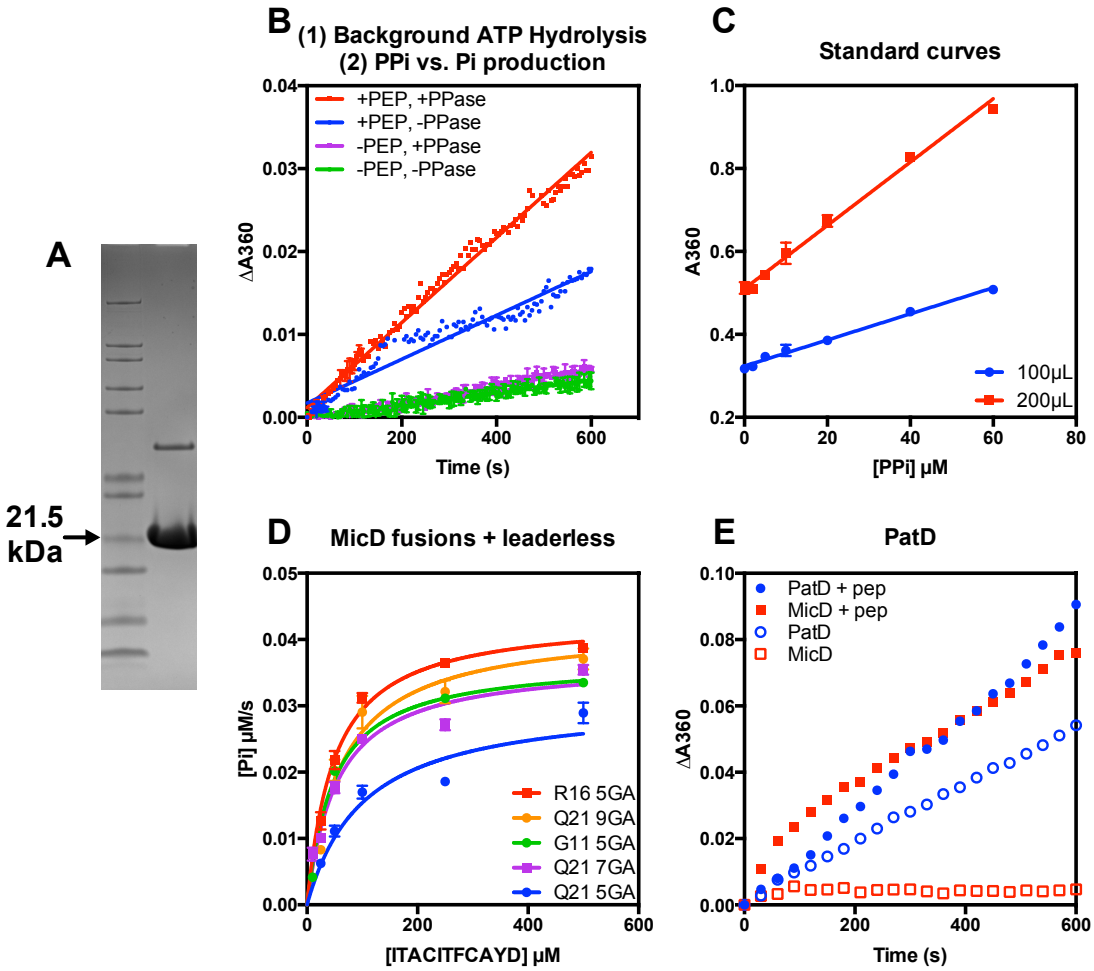

Figure 2.9: EnzChek Assays of MicD, PatD and fusion enzyme. (A) SDS-PAGE showing purified *M. tuberculosis* pyrophosphatase; (B) A360 readings of MicD R16 5GA (5 µM) reaction with 10 mM ATP and when specified, 100 µM ITACITFCAYD, in the presence of absence of 10 µM *Mt*PPase. Data was fitted to a linear regression. (C) Standard curves correlating changes in A360 to the concentration of inorganic pyrophosphate; (D) Initial rates of phosphate release plotted against substrate (ITACITFCAYD) concentrations, shown for each MicD fusion enzymes (5 µM Data was fitted in GraphPad Prism using equation 2.1; (E) A360 readings following hydrolysis of ATP by PatD or MicD (8 µM) or their reactions with 100 µM PatE2K.

Table 2.4: Steady-State Kinetics of Fusion Enzymes. The units of k_{cat} and K_m are in s^{-1} and µM, respectively, while k_{cat}/K_m is shown in min^{-1} M^{-1}.

Enzyme	k_{cat}	K_m	k_{cat}/K_m
MicD G11 5GA	0.0074 ± 0.0002	45 ± 3.9	9839
MicD R16 5GA	0.0087 ± 0.0003	46 ± 5.2	11322
MicD Q21 5GA	0.0062 ± 0.0006	98 ± 27	3769
MicD Q21 7GA	0.0073 ± 0.0003	51 ± 8.3	8619
MicD Q21 9GA	0.0084 ± 0.0006	59 ± 14	8503

2.5 Discussion

The construction of PatD fusion and MicD fusions reiterated the power of structural biology and rational design that was demonstrated by LynD fusion [64]. Binding of the leader peptide of PatE to native LynD induces structural changes which switch the enzyme into its active form (Figure 1.8 on page 11), and the fusion enzyme is believed to have the partial leader (recognition sequence) bound to the enzyme at all times, making it constitutively active. Without being structurally characterised themselves, PatD and MicD (and to some extent, OscD) acquired activities after modification. Independent result from Sardar and colleagues showed that the fusion strategy is also applicable to TruD, which becomes active on leaderless peptides when the sequence RLTAGQLSSQLAELSEEALG was appended to its N-terminus with a tetra Gly-Ser linker [126]. Moreover, lantibiotic synthetases have had repeated success with this method of engineering, similar to cyanobactin heterocyclases. In addition to LctM and HalCE2, mentioned in the introduction, a recent study fusing NisA leader to NisB [152] has demonstrated full dehydration activity in the modified enzyme. Based on these results it is highly likely that the fusion method can be widely extended to other leader-dependent RiPP PTMEs.

This work represents the first example of modified heterocyclases that synthesise both thiazolines and oxazolines, which will greatly expand the scope of heterocycle-containing peptides. The reaction was shown to be optimal in a high strength alkaline buffer (100 mM Tris pH 8.0 or higher) with low salt (50 mM NaCl) and reducing agents (5 mM). The fused enzyme was robust and tolerates sequential changes to a great level. Piggybacking on the LynD fusion construct, the leader-derived sequence that offers the strongest activation of the enzyme was found to be R16-A37 of PatE, which includes but extends beyond the recognition element 'LAELSEEAL' (Figure 2.9D). MicD fusion was also found to prefer a longer linker, showing enhanced activity when 18 residues instead of 12 (number of linker residues in LynD fusion) were placed between the protein and the appended sequence (Figure 2.9D). These optimisations led to a highly active biocatalyst that converts over 80 percent of the leaderless substrate into the fully heterocyclized product (Figure 2.6A). Compared to the established method [87] of heterocyclizing biologically produced, full-

length peptides, and then cleaving two thirds of the peptide before further modifications, vast amounts of material, enzyme and time can be saved by adopting the fusion enzyme, which uses short substrates that contain only the core sequence and the macrocyclase recognition signature. Having circumvented the need for a leader sequence on its peptide, MicD fusion can also be employed to process chemically synthesised peptides, which means unnatural amino acids and non-amino acids may be incorporated into the compound. This semi-synthetic approach was adopted by Oueis *et al.* [94, 95], resulting in the production of cyclo-peptides containing non-proteinogenic chemical groups as well as thiazoline. There is no reason why the heterocyclized linear intermediates from these studies cannot be subjected to other types of further modification such as prenylation, instead of being macrocyclized. Indeed, unnatural precursor peptides have been made with chimeric leaders that combine multiple recognition elements into the same leader [30], which was then reacted with multiple PTMEs, resulting in novel products that contain thiaozline and lanthionine, thiazoline and dehydroalanine, and so on, proving that heterocycles can co-exist with many other interesting chemical moieties. Moreover, in addition to the enhanced activity against a broadened substrate scope, the yield of fusion enzymes was significantly higher compared to native MicD; in the case of MicD R16 5GA, it was more than tripled (Table 2.3). Endowed with all the above stated benefits, MicD fusion makes a powerful piece of apparatus for the biological and chemobiological synthesis of natural and unnatural products.

An interesting observation was that MicD native (20 µM) heterocyclized a considerable fraction of the leaderless peptide ITACITACAYDGE twice (Figure 2.8C), while it could only turnover ITACITFCAYDG once in the same time frame (Figure 2.8B).The apparent explanation would be that AC heterocycles are easier/faster to form than FC heterocycles, probably due to the conformational restraints within the active site of the enzymes. This would however disagree with the observations made of TruD, which heterocyclizes all C8 cysteines before moving onto the C4 cysteines [65]. In addition, as you will read in chapter 3, MicD also cyclises FC first when reacted with ITACITFC, both when it is part of full-length PatE or a leaderless peptide. It is likely that conformational changes (such as a kink in the backbone) following the formation of thiazoline makes the singly heterocyclized peptide a less favourable substrate than its unmodified counterpart, and the FC thiazoline may cause more disruption than the AC heterocycle in this particular instance. Unlike MicD and TruD, LynD is an intrinsically disordered enzyme which shows no particular bias towards C4 or C8 on the full-length substrate, nor does LynD fusion has a preference for either cysteine when reacted with ITACITFCAYDG (chapter 3). In a previous study native LynD (5 µM) was reported to catalyse one residue only, when reacted with ITACITFCAYDG [64]. When a higher concentration of LynD (50 µM) were incubated with 50 µM ITACITACAYDGE and ITACITFCAYDG for 16 h, both reactions yielded a small amount of doubly heterocyclized peptides, with the latter peptide appearing to have undergone more turnovers (Figure S5, Appendix B) - contrary to how

MicD processes these peptides (Figure 2.8B and C). Although fully characterising the core sequence preferences of heterocyclases are beyond the scope of this chapter, it seems to vary between individual enzymes, and this variation could be exploited to generate diverse compounds. With regard to the leader peptide, the observations here showed that it is not essential for the heterocyclization of the second cysteine, unlike previously thought [64]; it does nevertheless accelerates the reaction significantly. The role of leader peptide in heterocyclization will be further explored in chapter 3.

Chapter 3

Role of Leader Peptide on the Rate and Order of Heterocyclization

3.1 Summary

RiPP PTMEs-catalysed precursor modification may follow various directionalities and orders of reaction. The cyanobactin heterocyclase TruD transfroms cysteines to thiazolines on the patellamide D core sequence ITACITFC in a C-to-N order. In this chapter, the directionality of heterocyclization when threonines are also catalysed has been determined using the homologue MicD, to be C8-C4-T2-T6 (as of ITACITFC). When the leader peptide is removed from the substrate, the reaction order was partially abolished and becomes C8-C4-T2/T6-T6/T2. In contrast, LynD, which heterocyclizes only cysteines, catalyses C8 and C4 without preference even when the leader peptide in present *in cis* to the core peptide.

3.2 Introduction

Precursor peptides of RiPPs may have multiple sites where modifications by a certain tailoring enzyme take place, and knowledge of reaction order with regard to those residues enhances the understanding of that enzyme's mechanism. From a biotechnological perspective, establishing the order of reaction could provide insight on how to capture specific reaction intermediates, which can be redirected to alternative modifications. PTME modifications could follow a particular direction (Figure 3.1A and B). Lantibiotic synthetases (classes I and II) often follow an overall N-to-C order in dehydrating serines

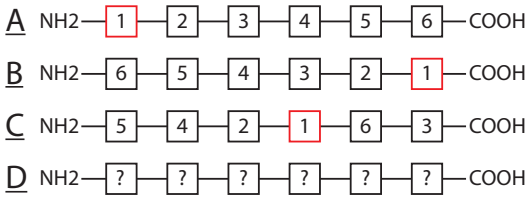

Figure 3.1: Schematic showing different reaction orders of RiPP modifications. (A) N- to C-terminal directionality. (B) C-to-N directionality. (C) Non-directional but ordered reaction. (D) Disordered mechanism.

and threonines to 2,3-didehydroalanine (Dha) and 2,3-didehydrobutyrine (Dhb) [89, 153], respectively, although this directionality was not strict, in the sense that residues in close proximity (such as S11 and T9 in LctA) could be catalysed in a random order, so long as both are processed before residues further C-terminal (e.g. S18 and S/T24 in LctA) [88]. In contrast, a class III enzyme LabKC was reported to operate in reverse, dehydrating residues S13-S10-S4-S1 [154]. Variation of directionality in related enzymes was also seen in azol(in)e synthetases involved in the biosynthesis of linear azol(in)e-containing peptides (LAPs) such McbBCD [155], which process in an overall N-to-C order and BalhBCD, which catalyses the majority of residues in a C- to the N-terminal sequence [68]. Further, non-directional but ordered reactions (Figure 3.1C) were have been reported, for example in the lantibiotic synthetase ProcM [156]. Leader peptide appeared to have an important role in maintaining the directionality and order of modification. Whereas native LctA sustains T9/S11 dehydration before S18/T24 dehydration, removal of leader from the substrate results in a change of order, where it becomes S11-T9-T24-S18, S11-T24-T9-S18, or T24-S11-T9-S18 [88]. Similarly, although the BalhBCD complex, which contains a leader recognition domain in the C protein, installs thiazolines on BalhA in a C-to-N direction, when the isolated oxidase BalhB (in the cited experiment, its homologue BcerB) is reacted pre-heterocyclized substrate, a cysteine located the middle of the sequence was transformed before the ones on both sides [113].

Cyanobactins PTMEs such as heterocyclase, oxidase and prenylase are capable of catalysing multiple residues (reviewed in chapter 1), yet little is known about the reaction order of these enzymes. The only study was that of TruD, a heterocyclase from the trunkamide pathway [65]. When presented with the patellamide D (Figure 3.2A) core sequence ITACITFC (as part of a precursor peptide containing leader), TruD transforms of cysteines to thiazolines in a defined order: the C-terminal cysteine (C51 of full sequence, C8 of core) first; it is only after all C51 has been catalysed that the enzyme starts to heterocyclize the other cysteine (C47, C4 of core). This means the enzyme works in a distributive manner - its dissociates from one substrate molecule and rebinds another without having catalysed all modifiable residues on the first molecule - with a much stronger preference for C51. Leader peptide plays a crucial role in the progression of

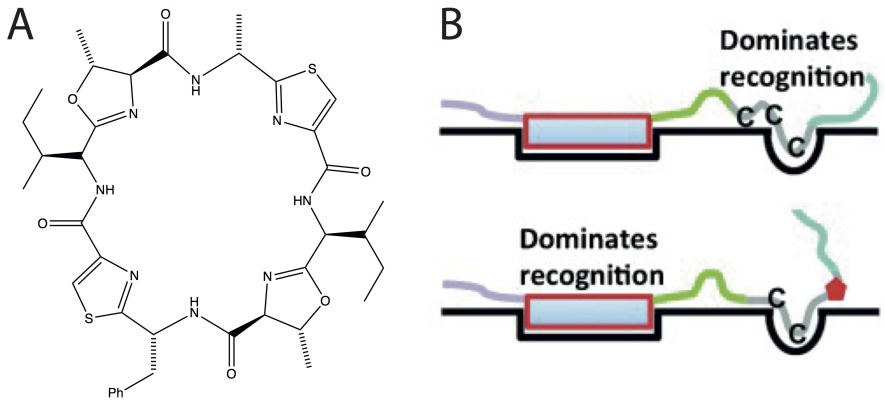

Figure 3.2: (A) Structure of Patellamide D. (B) Mechanism for heterocyclase substrate binding proposed by Koehnke and colleagues [65]. Adapted from [65].

the reaction, as when leader is removed from the substrate, C51 alone, but not C47 can be processed by TruD [65]. The authors proposed that (Figure 3.2B) the core peptide of linear PatE binds to the active site of TruD, while the leader peptide plays an auxiliary role in binding; once C51 is heterocyclized, the affinity for the core peptide lowers significantly, and the binding between leader and the protein now dominates, which keeps the core peptide in spatial proximity with the active site, leading to the catalysis of the second cysteine. However, it is not clear why C51 is favoured over C47 in the first place. It could be that leader binding creates an anchor point for PatE; as C47 is closer to the anchor, placing it in the active site could be structurally more strained than when C51 is positioned there. A further study carried using LynD [64], a homologue from the aestuaramides pathway, proved that leader peptide makes specific contacts with the enzyme, whilst the lack of electron density for the core region of peptide in the crystal structure of LynD suggests that the region is likely flexible. Reaction rates of LynD were compared when full-length substrate with *in-cis* leader was used, and when the leader peptide was added *in trans* to the core peptide or fused to the enzyme. The latter two scenarios led to a moderate reduction in the rate of the first heterocyclization and a more pronounced decrease in that of the second. Although this appears to corroborate core-dominated binding postulated from observations of TruD, It is worth pointing out that the C51-C47 order was not tested in the LynD study. On a side note, it seems contradictory that the core binds 'strongly' and is simultaneously flexible, and the leader may well be more important than previously thought in the binding of the unheterocyclized peptide.

In addition to thiazolines, which TruD and LynD are capable of producing, patellamide D (Figure 3.2) contains two methyloxazolines derived from threonines, and the reaction order when they are also catalysed is unknown. In this chapter, the full order of four heterocyclization events by the enzyme MicD was investigated in the context of the patellamide D core sequence ITACITFC, using a range of biochemical and biophysical techniques. The role of leader peptide in determining reaction order was then investigated

with the help of fusion enzymes created earlier in this project (see chapter 2). In addition, the reaction order of cysteine heterocyclization was revisited using the enzyme LynD. Lastly, kinetic analysis was carried out to further characterise the role of leader peptide. Together, these results revealed the complex interplay between chemoselectivity and spatial arrangement in the activity of cyanobactin heterocyclases.

3.3 Materials and Methods

3.3.1 Overexpression and Purification of Enzymes

LynD, LynD fusion, MicD and MicD fusion were overexpressed in *E. coli* BL21 (DE3) cells and purified through immobilised metal affinity chromatography and size-exclusion chromatography, as described in Section 2.3.5.

3.3.2 Cloning, Expression and Purification of PatE3KK

PatE3KK (MDEEN ILPQQ GQPVI ELTAG QLSSQ LAELS EEALG DAGLE ASKIT ACITF CAYDG EKLEH HHHHH) was created by introducing a lysine residue C-terminal to of E56 of PatE3 (MDEEN ILPQQ GQPVI ELTAG QLSSQ LAELS EEALG DAGLE ASKIT ACITF CAYDG ELEHH HHHH) via site-directed mutagenesis, following a protocol developed by Liu and Naismith [146]. pBMS-PatE3, supplied by Dr. Wael Houssen (University of Aberdeen), was used as a template and the primers 5'-CTGCATCACTTTTTGCGCTTAT GATGGTGAAAAACTCGAG-3' and 5'-GTGGTGGTGGTGGTGGTGCTCGAGTTTTTCACCATC-3' were used to introduce the lysine. PCR conditions are found in Section 2.3.4 on page 34. The mutated peptide was overexpressed and purified as described in Section 2.3.7 on page 37.

3.3.3 Preparation of Labelled Substrates

^{15}N-labelled PatE2K (MDKKN ILPQQ GQPVI RLTAG QLSSQ LAELS EEALD GAGLE ASKIT ACITF CAYDG ELEHH HHHH) was prepared as follows: A 10mL overnight culture of *E. coli* BL21 (DE3) cells carrying pBMS-PatE2K was inoculated into 1L M9 minimal medium [157] supplemented with 1 g ^{15}N-labelled NH_4Cl (Sigma), 0.4% (w/v) glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 100 µg/ml ampicillin and 0.17% (w/v) yeast nitrogen base. For ^{13}C, ^{15}N-labelled PatE2K, glucose was replaced with $^{13}C_6$-Glucose (Cortecnet). Cells were grown at 37 °C until OD_{600nm} reached 0.5, induced with 0.5 mM IPTG, and then incubated at 30 °C for 16 h. Cells were harvested and the protein was purified from cell lysates in the same way as non-labelled PatE (described in Section 2.3.7).

^{15}N-labelled PatE2K-2het was created by reacting ^{15}N PatE2K (100 µM) with LynD (5 µM) in 100 mM Tris pH 8 supplemented with 150 mM NaCl, 10 mM $MgCl_2$, 10 mM ATP and 5 mM DTT for 16 h at 25 °C. The heterocyclized peptide was purified by size exclusion chromatography on a 16/600 Superdex 75 pg column (GE Healthcare) in 20 mM Tris pH 8, 150 mM NaCl.

^{15}N-labelled PatE2K-2het-tryp was made by digesting ^{15}N PatE2K-2het with trypsin at 100:1 mass ratio at room temperature for 16 h. The reaction mixed was subjected to a nickel-NTA column (GE Healthcare) equilibrated with 20 mM Tris pH 8, 150 mM NaCl, and eluted with 20 mM Tris pH 8, 150 mM NaCl and 250 mM imidazole. The eluate (peptide expected to be 2468 Da) was dialysed into 10 mM Tris-HCl, pH 7.4, 150 mM NaCl using a dialysis cassette with 2K MWCO.

3.3.4 Nuclear Magnetic Resonance Spectroscopy

Three sets of NMR experiments were performed: (1) The reaction between 2 mM universally labelled (u)-^{13}C, ^{15}N-PatE2K and MicD, for sequential assignment as well as reaction monitoring. (2) The reaction between 100 µM universally labelled (u)-^{15}N PatE2K-2het and MicD. (3) The reaction between 100 µM u-^{15}N PatE2K-2het-tryp and MicD fusion (R16 5GA). All sets of experiments were performed at 20 °C on a Bruker Ascend 700 MHz spectrometer equipped with a Prodigy TCI probe, with the assistance of Dr Uli Schwarz-Linek. The instrument was controlled by Topspin (Bruker).

For (1), u-^{13}C, ^{15}N-PatE2K was concentrated to 2 mM and exchanged to a buffer containing 50 mM HEPES pH 7.4 supplemented with 150 mM NaCl, 15 mM ATP, 10 mM $MgCl_2$, 5 mM DTT, 0.02% sodium azide and 5% D2O. ^{1}H, ^{15}N-HSQC (heteronuclear single quantum coherence spectroscopy) spectra using a standard Bruker pulse sequence incorporating water flip-back and PEP water suppression at 2028 x 128 points and a digital resolution of 12.3 Hz and 31.0 Hz ^{1}H and ^{15}N dimension were recorded after 15 min, 30 min, 1 h and at each hour afterwards until 66 h after the start of the reaction. The reaction halted at this stage with the peptide in an intermediate state. ^{1}H, ^{15}N-HSQC spectra were recorded every hour. HNCACB and CBCA(CO)NH spectra were recorded using standard Bruker pulse sequences with 1536 x 50 x 104 points and digital resolutions of 12.8 Hz, 59.6 Hz, and 257 Hz for the ^{1}H, ^{15}N and ^{13}C dimension, respectively, for the staring material and the final product.

Assignment of PatE2K at 20 °C was performed by Dr Ona Miller and Dr Uli Schwarz-Linek. Assignment of PatE2K-MicD reaction intermediate at 66 h and the final product was conducted with the assistance of Dr Uli Schwarz-Linek.

All spectra were processed with NMRPipe [158] and analysed with CCPN Analysis 2

[159].

For (2), 100 µM u-^{15}N-PatE2K-2het was reacted with 5 µM MicD at 20 °C and the reaction was monitored by taking ^{1}H, ^{15}N-HSQC spectra every 30 min for a total of 16 h. Four transients were recorded at 2048x110 points with spectral resolution of 12.3 Hz and 36.1 Hz for the ^{1}H and ^{15}N dimensions, respectively.

For (3), 100 µM u-^{15}N-PatE2K-2het-tryp was reacted with 5 µM MicD fusion (R16 5GA) at 20 °C.^{1}H, ^{15}N-HSQC spectra were collected before and 16 h after the reaction; six transients were recorded at 1536x80 points with spectral resolution of 12.8 Hz and 26.6 Hz for the ^{1}H and ^{15}N dimensions, respectively.

Spectra collected from experiments (2) and (3) were processed with Bruker Topspin and analysed with CCPN Analysis 2 [159].

3.3.5 Alkylation of Free Cysteines

PatE2K (100 µM) was reacted with MicD (5 µM) for 1.5 h, 3 h and 16 h in 20 mM Tris pH 8.0, 150 mM NaCl, 10 $MgCl_2$ and 10 mM ATP. A sample was removed from the reaction and quenched with 50 mM EDTA after each time point. 5 mM alkylating agent iodoacetamide (IAA) was added to each quenched sample, which was then kept away from light for 30 min at room temperature and subjected to MALDI mass spectrometry.

ITACITFCAYD (100 µM) was reacted with MicD (5 µM) for 1 h, 2 h, 3h and 16 h in 20 mM Tris pH 8.0, 150 mM NaCl, 10 $MgCl_2$, 10 mM ATP and 5 mM DTT. The reaction was quenched as above but was diluted 1 in 10 in water before the addition of 5 mM IAA. The alkylated samples were incubated at room temperature for 30 min away from light and subjected to MALDI mass spectrometry.

3.3.6 Mass Spectrometry

Unless specified, mass determination was obtained through MALDI mass spectrometry on a 4800 MALDI TOF/TOF Analyzer (ABSciex).

For reaction order determination, tandem mass spectrometry (MSMS) data were collected. MicD and LynD (5 µM each) were reacted with PatE3KK (100 µM) individuallyfor 1 h and 16 h at 20 °C in heterocyclization buffer (100 mM Tris pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM ATP and 5 mM DTT). The reaction was stopped by adding EDTA to the final concentration of 50 mM. Trypsin was added at 1:100 mass ratio followed by incubation at 20 °C for 16 h. MicD fusion and LynD fusion (5 µM each) were incubated with ITACITFCAYDG (100 µM) individually in heterocyclization buffer for 1

h or 16 h, and quenched with 50 mM EDTA. The reaction mixtures were subjected to high performance liquid chromatography and tandem MS spectrometry (HPLC-MSMS) on a AB SCIEX Triple TOF 5600 system equipped with Eksignet nanoLC. The samples were first desalted by being entered into a 0.1 ml/min flow of 98% H_2O, 2% acetonitrile (ACN) and 0.05% trifluoroacetic acid (TFA), and washed through a Thermo Pepmap 20 mm x 0.075 mm column (trap column) for 5 min. The trap column was then connected to an analytical column (150 mm x 0.075 mm) and both were equilibrated in 98% H_2O, 2% ACN and 0.1% formic acid (FA). Peptides were eluted and separated by switching the solvent towards 98% ACN, 2% H_2O and 0.1% FA in a linear gradient over the course of 6 min, with a flow rate of 300 nl/min, and the eluate was passed through a first mass spectrometer. From there, the most intense peaks within the 400-1250 m/z region were automatically selected in the Information Dependent Acquisition (IDA) mode and directed to Collision Induced Fragmentation (CID) followed by a second MS scan in the region of 95-1800 m/z. Alternatively, Product Ion Scan (PIS) mode was employed where specified masses were entered into the programme prior to experiment so that they are selected for fragmentation and MSMS. Data were collected by Dr Sally Shirran, Dr Catherine Botting or Dr Silvia Synowsky from the University of St Andrews mass spectrometry facility.

3.3.7 Enzyme Kinetics

Reaction rates of MicD (1 µM) and MicD fusion (MicD Q21 5GA, 5 µM) enzymes were measure using a commercially available phosphate assay as described in Section 2.3.11. Reaction rates of LynD (0.4 µM) and LynD fusion (5 µM) was measured using the same assay, but the reaction mix contained 1 mM ATP and 1 mM $MgCl_2$, instead of 10 mM of each. MicD and LynD were reacted with PatE2K, while MicD fusion and LynD fusion were reacted with ITACITFCAYD and ITACITFCAYDGE, respectively. Up to 500 µM leaderless peptides were used due to solubility issues. Experiments were carried out in triplicates and fitted to the quadratic equation (Eq 2.2). Reaction with higher concentrations of PatE2K were carried out using 1 µM MicD (triplicates) or 2 µM LynD (no replicates) and fitted to the substrate inhibition equation (eq 3.1) in Prism. v_{int}, initial rate; $[E_T]$, total enzyme concentration; k_{cat}, catalytic constant; $[S_T]$, total substrate concentration; K_m, Michaelis constant; K_i, inhibition constant.

$$\frac{v_{int}}{[E_T]} = \frac{k_{cat}[S_T]}{K_m + [S_T](1 + \frac{[S_T]}{K_i})} \tag{3.1}$$

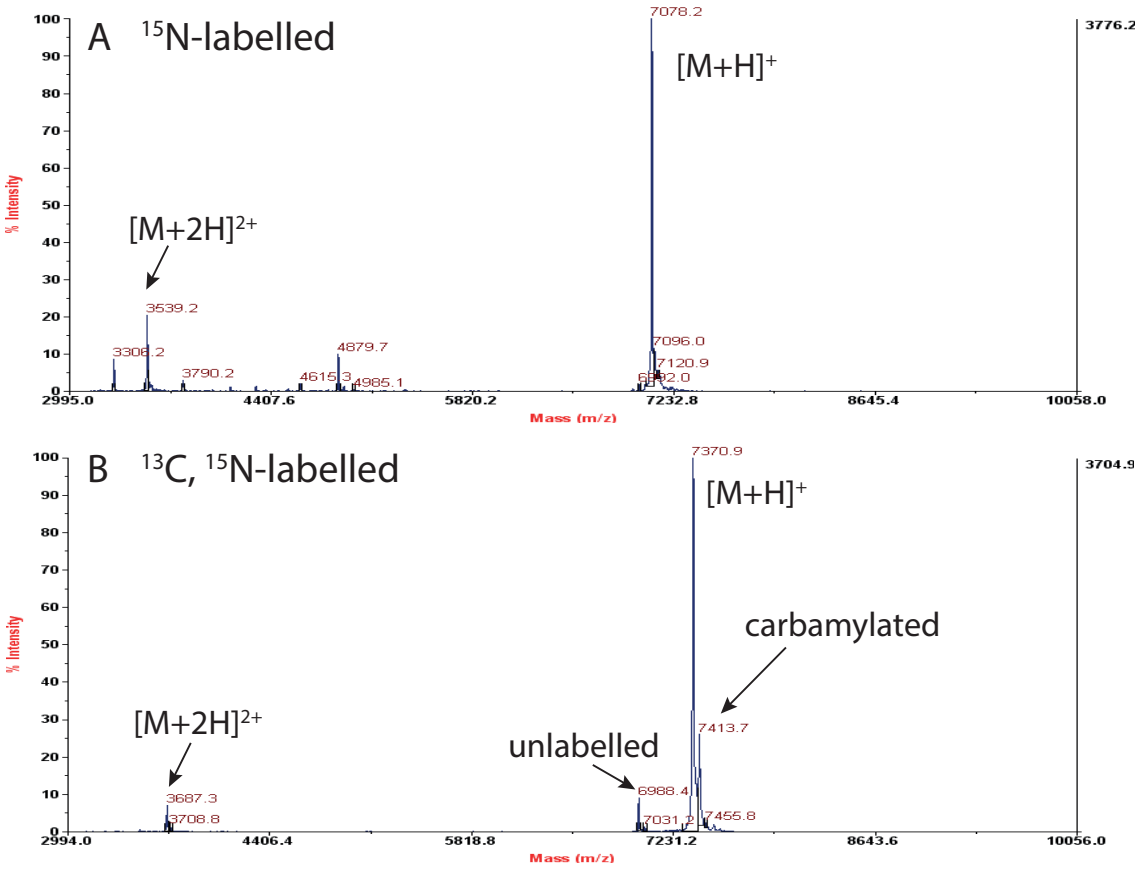

Figure 3.3: MALDI analyses of ^{15}N- (A) and ^{13}C, ^{15}N-labelled PatE2K (B).

3.4 Results

3.4.1 Production of Labelled Substrates for NMR

Singly and doubly labelled PatE2K were purified and subjected to mass spectrometry (MS). ^{15}N-labelled PatE2K has an expected mass ($[M+H]^+$) of 7078 Da, and the MS result (Figure 3.3A) showed complete labelling. In contrast, the observed mass of the ^{13}C, ^{15}N-labelled peptide was 7370 Da (Figure 3.3B), 10 Da smaller than the calculated value of 7380 Da, indicating that carbon labelling has a lower efficiency. A small proportion (less than 10%) of the peptide was unlabelled, as shown by the 6988 Da peak. A 43 Da adduct was also observed, which likely arose from carbamylation of lysines during sonication (lysis buffer contained 8M urea). As the unlabelled material will not show in ^{1}H, ^{15}N-HSQC spectra, and the lysine side chains in the unfolded peptide will be too far away from the polypeptide backbone to affect mainchain resonances, this sample was deemed to be of acceptable quality for NMR experiments.

Doubly heterocyclized ^{15}N-PatE2K (^{15}N-PatE2K-2het) was obtained by reacting ^{15}N-

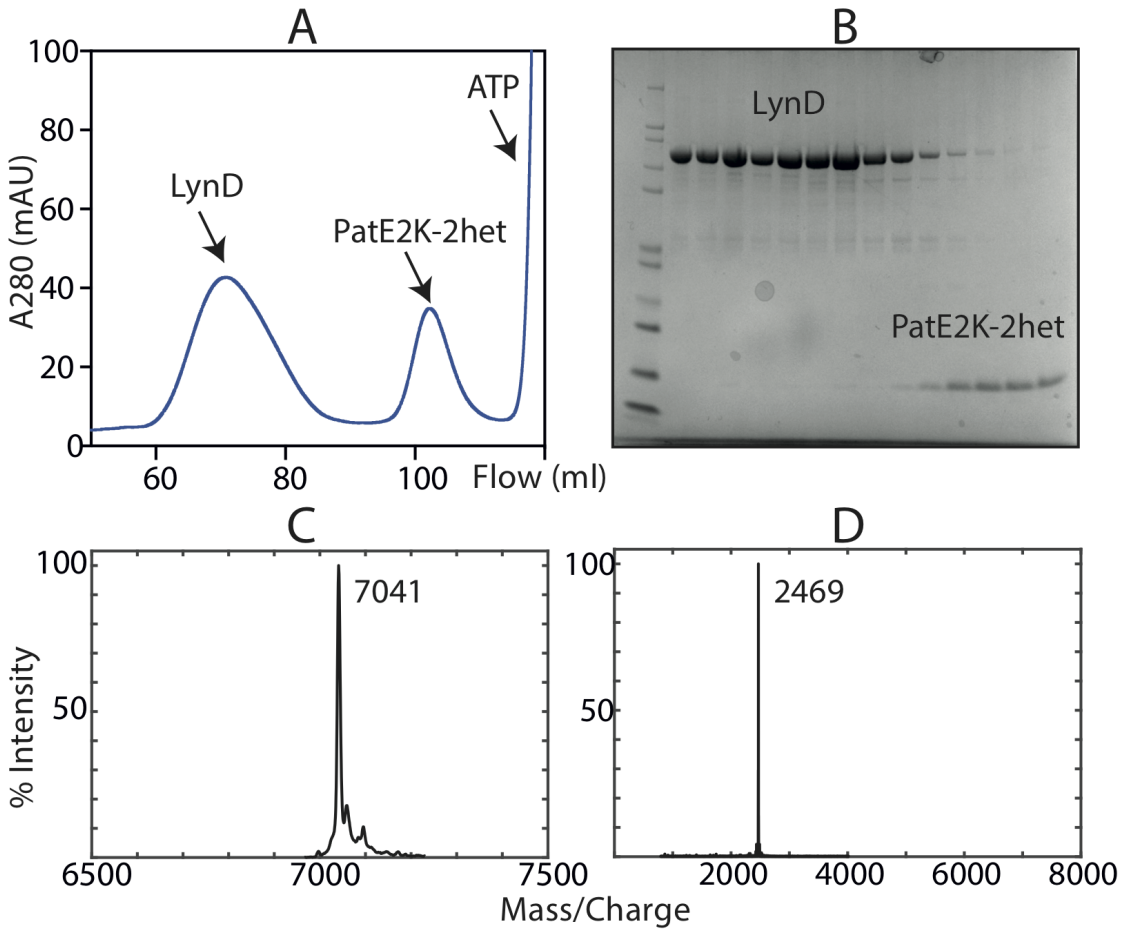

Figure 3.4: Preparation of heterocyclized full-length and leaderless peptides. (A) Separation of LynD and PatE2K on Superdex 75 column. (B) SDS-PAGE of elution fractions from (A). (C) MALDI spectrum of purified ^{15}N-PatE2K-2het. Expected 7041 Da. (D) MALDI spectrum of purified ^{15}N-PatE2K-2het-tryp showing the 0-8000 m/z range. Expected 2468 Da.

PatE2K with LynD. The modified peptide was then separated from LynD by size exclusion chromatography (Figure 3.4A and B). MS analysis confirmed that two heterocycles were formed (Figure 3.4C). Leaderless, doubly heterocyclized ^{15}N-PatE2K (^{15}N-PatE2K-2het-tryp) was obtained by trypsin cleavage of the N-terminal leader of ^{15}N-PatE2K-2het, while the remaining C-terminal His-tag (sequence: ITACITFCAYDGELEHHHHHH) was used for purification by affinity chromatography. Again, MALDI-MS analysis confirmed that proteolysis of the leader peptide has been complete (Figure 3.4C).

3.4.2 PatE3KK Was Used to Produce ITACITFCAYDGEK

PatE3KK was designed such that the leader peptide (MDEENILPQQGQPVIELTAGQLSSQLAELSEEALDGAGLEASK) and the His-tag tail (ELHHHHHH) are removed from the core peptide (ITACITFCAYDGEK) upon

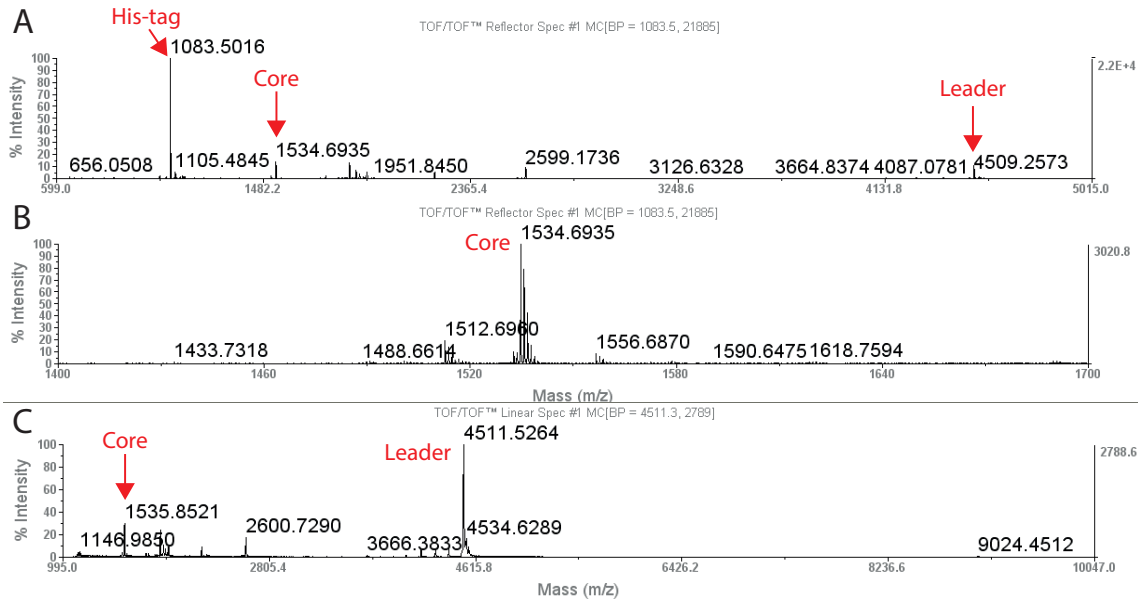

Figure 3.5: MALDI Spectra of trypsinated PatE3KK. (A) Ions in the range between 599 and 5015 Da in the reflector mode. (B) Zoom-ion of the region that contains the ion that corresponds to the core peptide. (C) Ions ranging from 995 to 10047 Da in linear mode, where the leader peptide ionises more prominently

trypsin cleavage. PatE3KK was used to react with native heterocyclases before being trypsinated to give the cleaved core peptide, which is highly similar to the leaderless substrates (such as ITACITFCAYDG) that were reacted with fusion heterocyclases, thus allowing the activities of native and fusion enzymes to be compared readily. Figure 3.5 shows that PatE3KK is digested into 3 segments after the addition of trypsin, as expected.

3.4.3 MicD Catalyses C51 First

PatE2K, which encodes the patellamide D core sequence ITACITFC, was used as a model to study the order of heterocyclization when cysteines and threonines are both transformed. It was known that TruD catalyses the C-terminal cysteine (C51/C8) before the internal one (C47/C4) [65], but with MicD the order could change. We reacted PatE2K with MicD and monitored the formation of heterocycles using nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS).

First, heteronuclear single quantum coherence spectroscopy (^{1}H, ^{15}N-HSQC) experiments were set up to track the reaction between unheterocyclized PatE2K and MicD. This technique reports on the chemical environments of different amide protons, giving distinct cross peaks for each residue in a polypeptide chain. Heterocyclization of a residue would result in the loss of its amide proton, and thus the disappearance of its cross peak

on the HSQC spectrum. Meanwhile, resonances from adjacent residues will also shift as a result of the chemical change. The identities of cross peaks are determined by recording CBCANH and CBCA(CO)NH spectra (Figure 3.6A and B), which are triple resonance (3D) experiments that establish links between the ^{1}H, ^{15}N and ^{13}C dimensions [160]. Each amino acid has its distinct pattern of Cα and Cβ chemical shifts (Figure 3.6C), and the carbon resonance profile (shown as strong peaks in the CBCANH spectrum, Figure 3.6D) at a specific ^{1}H x ^{15}N combination (a cross peak) reveals the residue type of that cross peak, for example, alanine. Moreover, these experiments allow assignment of residues to specific positions in a sequence (e.g. A46). As well as showing strong resonances for the carbons that belong to the same amino acid as an NH group, CBCANH shows weak signal of the carbons from the preceding residue, thus establishing sequential relationships between different cross peaks. CBCA(CO)NH, on the other hand, gives strong signals that correlates each NH group to the Cα and Cβ of its preceding residue, and can be used to complement CBCANH. Softwares such as Analysis [159] allow users to search for specific resonance values (Figure 3.6D). If serine is shown to be preceded by an alanine, the minor resonances from the serine's CBCANH spectrum can be used to search for the alanine's CBCANH spectrum, which will in turn unveil what precedes the alanine. After this process is repeated several times it is usually possible to obtain a segment (such as GLEAS) that is unique in the protein so that their residue numbers can be determined.

PatE2K gives a clustered spectrum characteristic of a natively unfolded protein (Figure 3.7C). After reacting with MicD for 4 h at 20 °C, the cross peak for C51 has disappeared (Figure 3.7A), consistent with the loss of its amide proton. Accompanying that was the shifting of cross peaks for I48, T49, F50, A52, Y53, D54 and G55 (Figure 3.7A and B), residues that flank C51. The closer to C51 a residue was, the more distance its cross peak travelled - F50 shifted by 0.26 and 3 ppm in the ^{1}H and ^{15}N dimensions, respectively, while T49 shifted only by 0.05 (^{1}H) and 0.46 (^{15}N) ppm (Figure 3.7A). These results confirmed that like TruD [65], MicD had a clear preference for C51 (C8 of core peptide).

After the first heterocyclization, the reaction rate decreased considerably, and the situation became more complicated. In the spectrum collected 66 h after the addition of enzyme, the peptide was seen in an ensemble of multiple intermediates states (Figure 3.8A). Cross peaks for C47, T45 and T49 have weakened but not disappeared, suggesting that heterocyclization of these residues have started, yet not completed. When compared to uniformly doubly heterocyclized (2-het) PatE2K (Figure 3.8B), which contains two thiazolines derived from C47 and C51, it was evident that singly heterocyclized (1-het) peptide still was part of the ensemble after 66 h of reaction with MicD, as T45, A46, C47, I48, T49, F50 and A50 in the 66 h spectrum (red) showed resonances that overlapped with those in the 4 h spectrum (blue). In the meanwhile, these residues also gave cross peaks that corresponded to the C47, C51-heterocyclic state of PatE2K (green). New peaks that were not observed in the either the uniformly 1-het (blue) or 2-het (green) species were seen

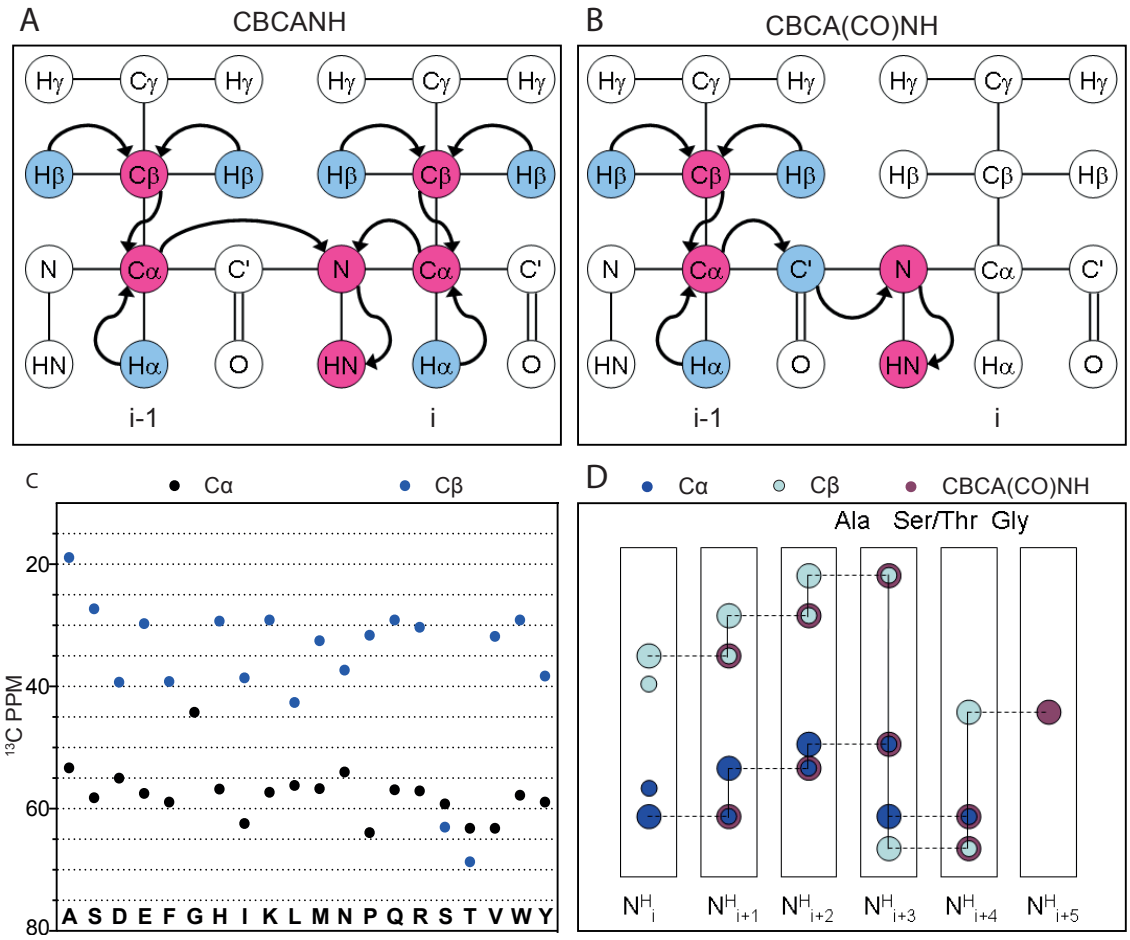

Figure 3.6: Triple resonance backbone assignment theory. (A) and (B) show how magnetisation is transferred between atoms in CBCANH and CBCA(CO)NH experiments, respectively [160]. (C) Typical ^{13}C chemical shifts of 20 amino acids at pH 7.5, 25°C, plotted using reference numbers given in [161]. (D) Schematic demonstration of sequential assignment in practice. (A), (B) and (D) were taken from protein-nmr.org.uk and adapted.

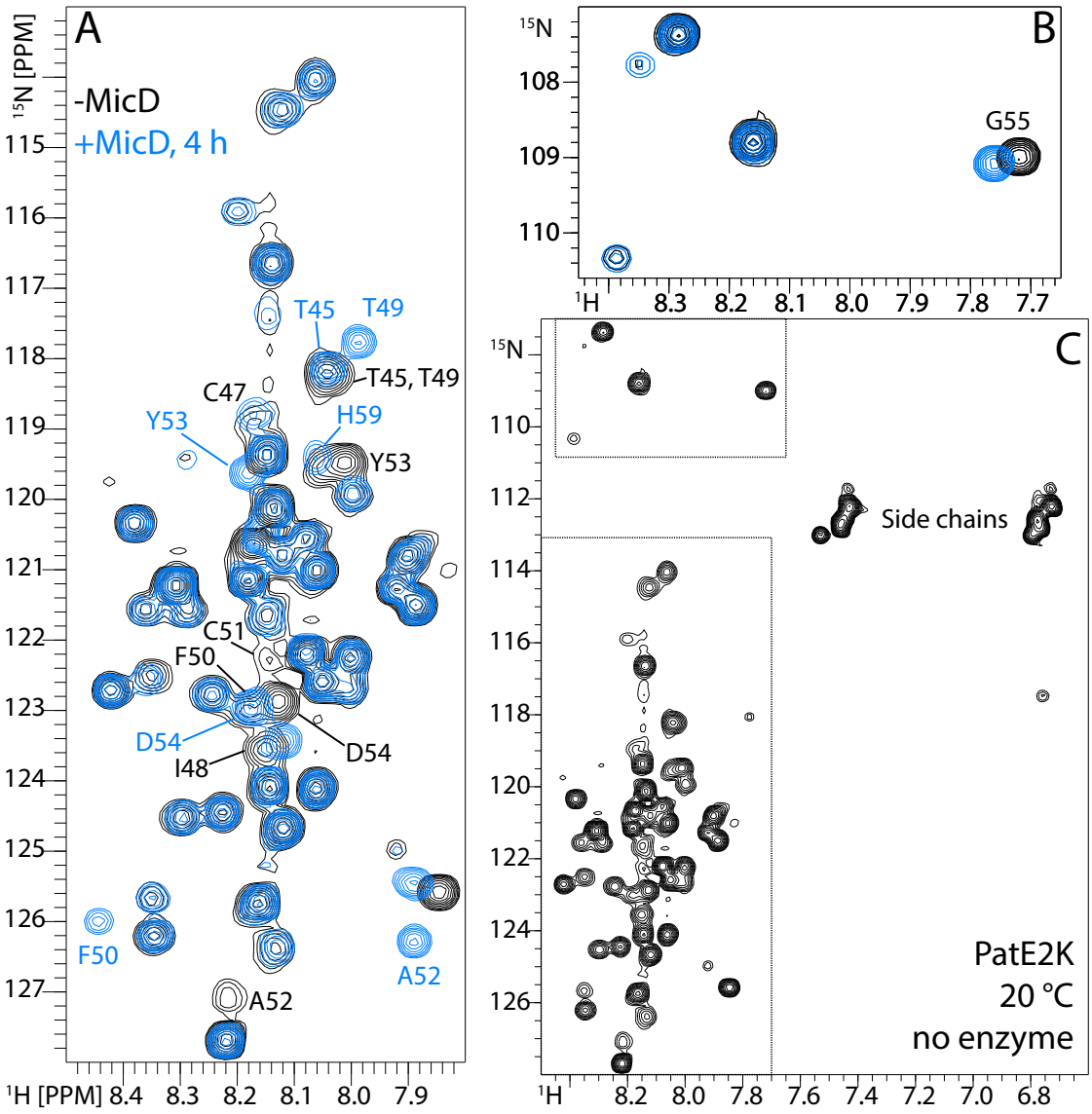

Figure 3.7: (A)(B) ^{1}H, ^{15}N-HSQC spectrum of PatE2K after 4 h of reaction with MicD (blue) overlaid with unreacted PatE2K (black). (C) Full spectrum of unreacted PatE2K at 20 °C. Dashed boxes highlight areas that are shown in (A) and (B)

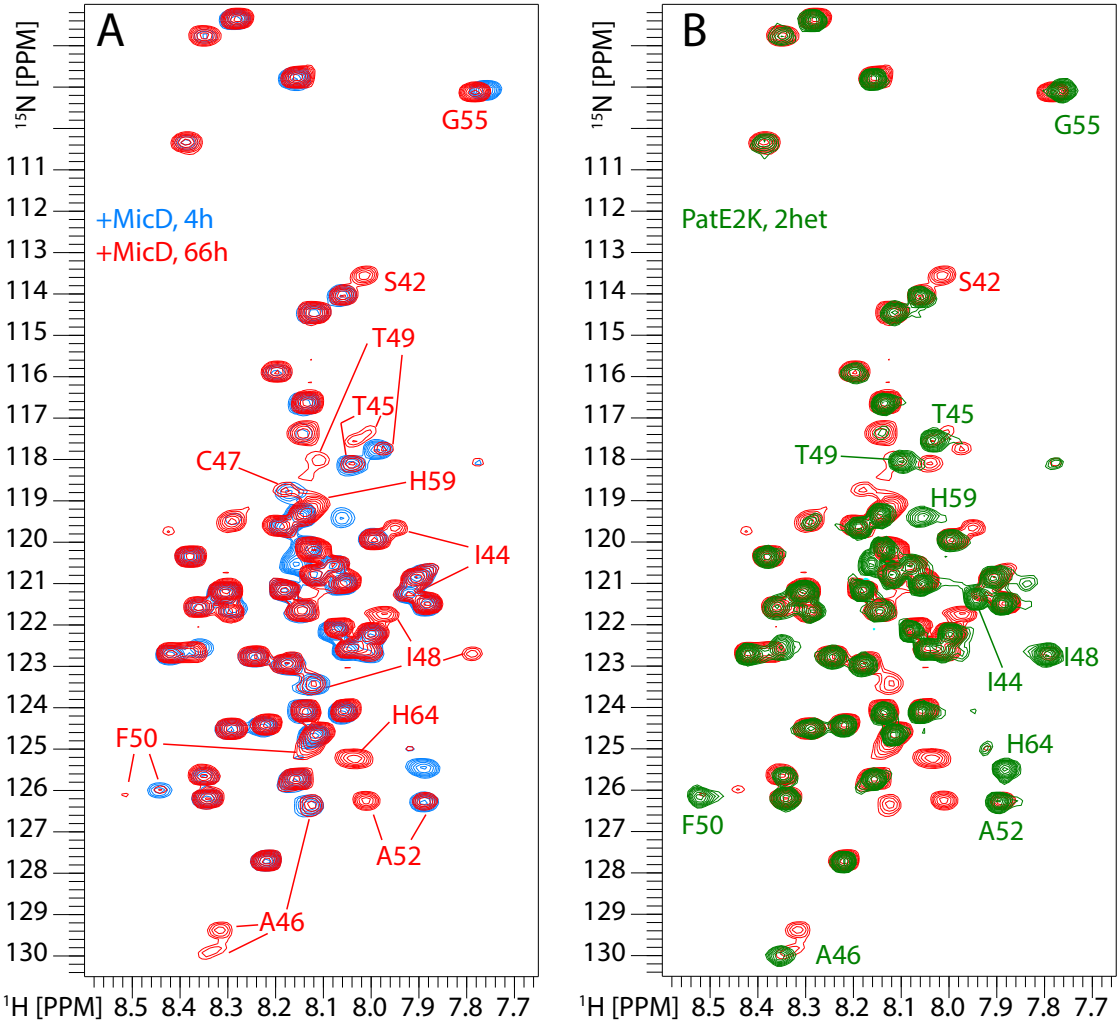

Figure 3.8: (A) Overlay of ^{1}H, ^{15}N-HSQC spectra of PatE2K reacted for 66 h (red) and 4 h (blue). (B) Overlay of ^{1}H, ^{15}N-HSQC spectra of PatE2K reacted for 66 h (red) and PatE2K with two C47 ad C51 heterocycles (green).

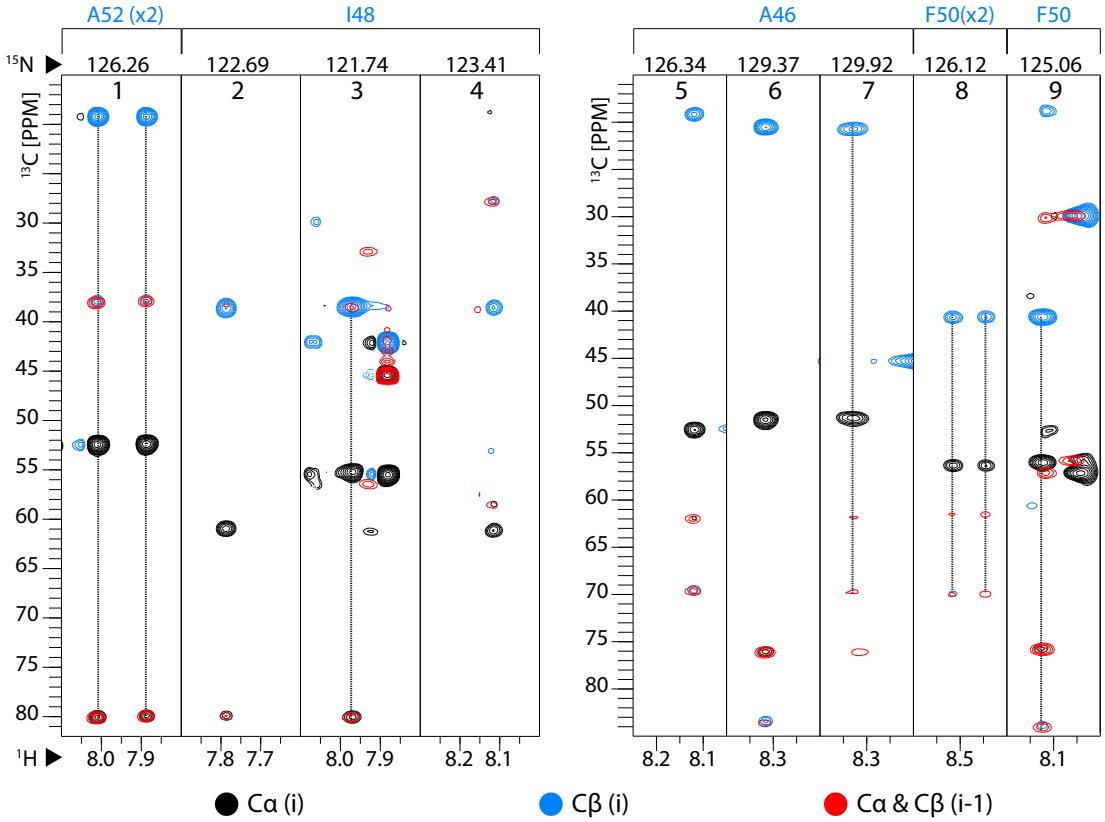

Figure 3.9: CBCANH (black and blue) and CBCA(CO)NH (red) spectra of key residues.

in the MicD-reacted sample (red) - S42 at [8.01, 113.55] (^{1}H and ^{15}N ppm, respectively), I44 at [7.95, 119.66], A46 at [8.32, 129.38], I48 at [7.97, 121.74], F50 at [8.14, 125.04] and A52 at [8.01, 126.25] - showing that additional chemical shifts have occurred as a result of threonine heterocyclizations. To conclude, MicD does not heterocyclize all the C47 residues before it moves on to T45 and T49, in contrast to how it catalyses C51. This however does not rule out an ordered reaction mechanism - if C47 heterocyclization precedes that of T45 and T49, but once C47 is catalysed latter transformation are carried out at a similar rate, changes corresponding to all three azoline formations will show on the NMR spectra.

In an attempt to infer the reaction order through further analysis, 3D spectra of residues directly C-terminal to the heterocyclization residues, namely A46, I48, F50, and A52, were used to determine if T45, C47, T49 and C51, respectively, have been catalysed. Heterocyclization change the carbon shifts of a residue drastically. Both A52 peaks from the red spectrum in Figure 3.8A are preceded by a thiazoline (Figure 3.9, strip 1), which gives a prominent Cα (i-1) shift at around 80 ppm, not typical of any amino acid (Figure 3.6C). The peak [7.89, 126.26] which matches A52 in PatE2K-2het (Figure 3.8B) probably formed first, upon C47 heterocyclization, and then shifted to [8.01, 126.25] when a threonine is cyclized, most likely T49 due to its relative proximity

to A52. Presuming both T49 and C47 cyclodehydration affect the chemical shifts of A52, a different intermediate/cross peak should have resulted if T49 is cyclized before C47. Similarly, I48, which follows C47, has two (out of three) peaks with CBCANH spectra that show carbon resonances of thiazoline (Figure 3.9, strips 2 and 3), one of which ([7.79, 122.70]) overlaps with I48 from the uniformly doubly heterocyclized PatE2K. In contrast, only one of the three A46 peaks ([8.32, 129.38], strip 6) and one of the three F50 peaks ([8.14, 125.04], strip 9) are preceded by a methyloxazoline, which gives Cα and Cβ shift values at around 76 and 84 ppm, respectively. Unless C47 heterocyclization has no effect on A46/F50 after T45/T49 is transformed, there should be additional peaks if C47 is sometimes heterocyclized after either threonine. While these observations point to C47 being catalysed before T45 and T49, more concrete evidence is needed to prove or disprove that, as well as whether there is an reaction order of methyloxazoline formation. Experiments described in the next two subsections were therefore carried out.

3.4.4 Cysteines are Heterocyclized Before Threonines

Iodoacetamide (IAA) is an alkylating agent that reacts with the thiol group of free cysteines and as a result adds 57 Da to the mass of the modified peptide/protein. Unmodified PatE2K, which contains two free cysteines, would acquire a +114 Da mass shift after alkylation. If one of the cysteines is heterocyclized the mass shift would be +57 Da, whereas if both are catalysed there would be no mass addition upon reaction with IAA. By alkylating modified PatE2K with different numbers of heterocycles, mass shifts could be used to determine how many free cysteines are left, and thereby to calculate how many of those heterocycles are thiazolines. Figure 3.10A shows PatE2K after 4 h of reaction with MicD. A mixed state with species corresponding to the starting material (SM), singly heterocyclized (1het), doubly heterocyclized (2het) and triply heterocyclized (3het) peptide was observed. Upon reaction with 5 mM IAA (Figure 3.10C), the 1het species made a mass shift of +57 Da, except for a residual level peak at the original mass, suggesting that the first heterocyclization was carried out on one of the cysteine residues. In contrast, masses of the 2het and 3 het species did not change, consistent with a scenario where both cysteines were no longer susceptible to alkylation. The majority of unheterocyclized peptide (6989 Da) was modified twice, although a noticeable fraction of it showed only one 57 Da mass adduct, which could be due to formation of disulphide-linked PatE dimers/aggregates which competed with IAA labelling, as DTT was not added prior to alkylation.

After 16 h of reaction with MicD (Figure 3.10B), a small fraction of the peptide (close to noise level) remained in the 1het state, which was modified by IAA like in the 1.5 h sample (Figure 3.10D). The vast majority of PatE2K has undergone two to four heterocyclizations (Figure 3.10B), giving mass over charge (m/z) peaks that did not

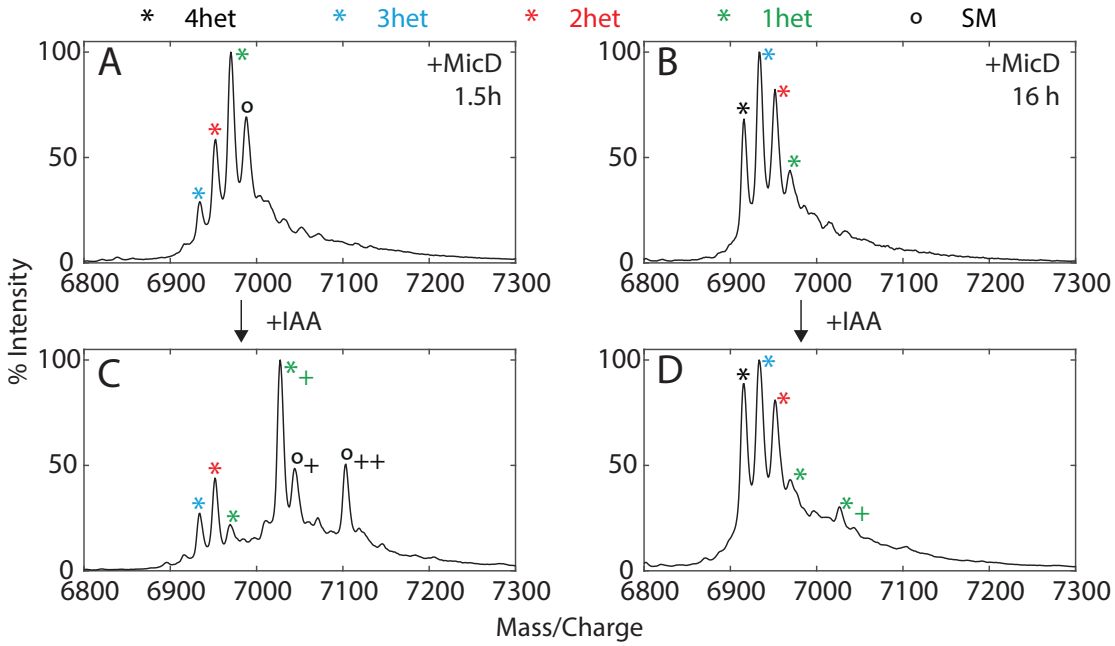

Figure 3.10: Alkylation of PatE2K Reacted with MicD. (A)(C), PatE2K reacted with MicD for 1.5 h. (B)(D), PatE2K reacted with MicD for 16 h. (B) and (D) were treated with 5mM iodoacetamide (IAA) l; 1het - 4het, peptide with one to four heterocycle(s); '+' and '++' signify mass shifts of +57 Da and +114 Da, respectively.

shift after IAA modification (Figure 3.10D). The high abundance of 2het-4het species consolidated the conclusion from Figure 3.10A and C that they do not contain free cysteines. It logically follows that the first two heterocyclization events take place at cysteine residues; in other words, cysteines are heterocyclized before threonines.

3.4.5 T45 is Heterocyclized before T49

After establishing that C51 and C47 are the first two residues to be heterocyclized, the reaction order between T45 and T49 was analysed. NMR spectroscopy was employed again, this time using doubly heterocyclized PatE2K (PatE2K-2het) as a starting material, with ^{1}H, ^{15}N-HSQC spectra collected every 30 min after the addition of MicD. The black spectrum in Figure 3.11 represents purified PatE2K-2het is equivalent to the green spectrum in Figure 3.8. Cross peaks corresponding to residues S42, K43, I44, A46, I48, F50, and A52 shifted as the peptide was transformed by the enzyme, notably only once each (Figure 3.11, blue and red spectra). After 8.5 h of reaction with MicD at 20 °C, resonance for K43, I44 and T45 disappeared, and that for S42 were significantly weakened. A46, I48, T49, F50 and A52 also reduced in intensity, albeit to a lesser extent. After 16 h of reaction all residues have completely shifted or disappeared except for a residual amount of A52. Quantitative analysis of peak volumes corroborated the rate differences (Figure 3.12). Although residues varied in their maximum intensity (Figure 3.12A), after

normalisation they formed two distinct clusters with regard to the rate of intensity decay (Figure 3.12B) - cross peaks S42, I44 and T45 diminished from their original position more rapidly than A46, I48, T48, A50 and F50. The fact that the A46 cross peak decreased at a similar rate to residues that are closer to T49 was unexpected, as it is closer to T45; also, the new A46 peak (Figure 3.11, blue and red spectra) has the same shift ([8.32, 129.38]) of the A46 peak in the Figure 3.8 (red spectrum), which is known via 3D experiments to be C-terminal to a oxazoline. Despite this oddity, as rates were consistent between S42, I44 and T45, as well as between T49, F50 and A52, these residues were used as indicators of T45 and T49 heterocyclization, respectively. Normalised peak volumes of the two sets of peaks were converted to peptide concentration, averaged and fitted to linear regression (Figure 3.12C) to obtain an approximation of reaction rate. The average rate of T45 heterocyclization, between the beginning of reaction and when signals of relevant cross peaks were not detectable, was 11± 0.5 µM/h, whilst that of T49 was 7 ± 0.3 µM/h. The results showed that MicD catalyses T45 faster than it does T49; nevertheless, it is not clear whether T49 heterocyclization takes place on the triply heterocyclized peptide as it is formed, or it occurs along side the T45 reaction at a slower rate. As discussed in section 3.4.3, the fact that multiple varied intermediates were not observed makes the former mechanism more plausible, but further experiments are required to validate this hypothesis.

It was then reasoned that as triply heterocyclized PatE could contain either T45 or T49 heterocycle, subjecting this peptide to tandem mass spectrometry (MSMS) would reveal whether the methyloxazoline preferentially formed at one position. Figure 3.13A gives an brief overview of the principles underlying MSMS: primary ions are separated by their mass (m/z) and those of a specific mass can be selectively directed to fragmentation. The separation and selection is particularly important for this study, as PatE tends to be in an ensemble state with multiple intermediates after reacting with MicD (Figure 3.8, Figure 3.10). Figure 3.13B shows the ions that commonly result from peptide fragmentation. In this experiment, the locations of heterocycles on the peptide were determined following two rules, using b ions and y ions, which arise from breakage of the peptide bond. First, if a particular fragment (e.g. ITA) exhibits a -18 Da shift compared to the calculated mass, that ion likely contains a heterocycle. Second, once a heterocycle forms, the amide bond N-terminal to the modified residue (i) no longer exists, and therefore the b_{i-1} and the y_{n-i+1} ions would disappear, where n equals the total number of residues in the peptide. A list of the mass calculations taken into account heterocyclization is shown in Table 3.2 and Table 3.3 for b ions and y ions of ITACITFCAYDGEK, respectively, with nomenclature of the heterocyclized intermediates/product given in Table 3.1.

PatE3KK was reacted with MicD and cleaved by trypsin to release the peptide ITACITFCAYDGEK. An total ion count (TIC) scan for the mass corresponding to the triply heterocyclized peptide (740.84 Da) gave three peaks between 11 and 13 min of run

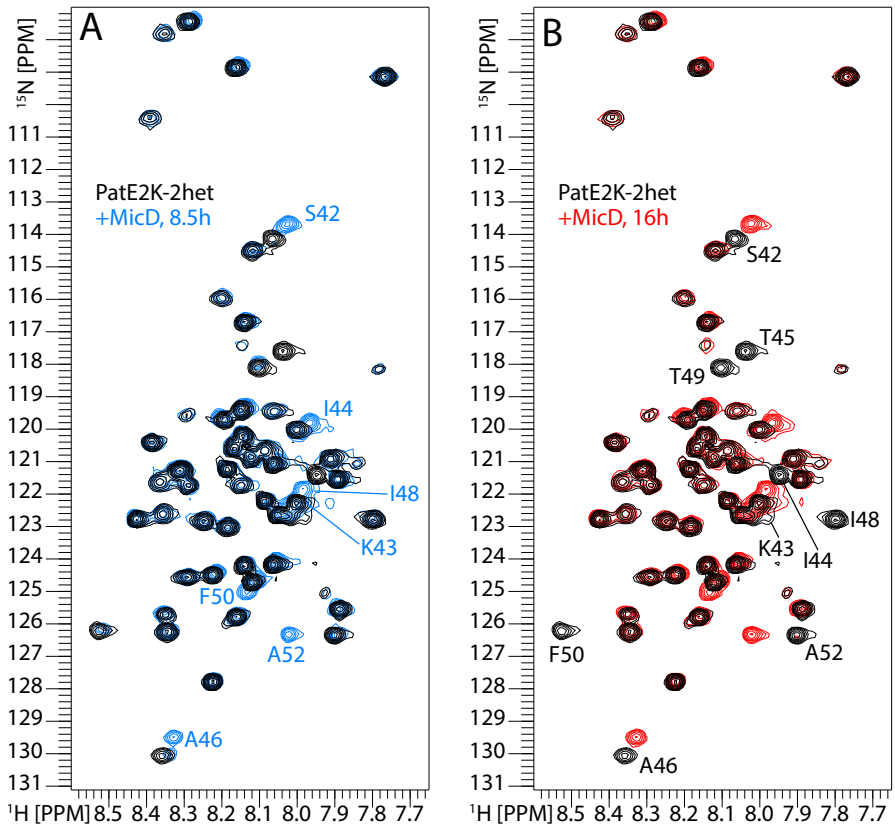

Figure 3.11: ^{1}H, ^{15}N-HSQC Spectra of PatE2K-2het Reacted with MicD. PatE2K-2het (100 µM) was reacted with MicD (5 µM) at 20 °C for 16 h. The black spectrum (A and B) shows resonances from the unreacted peptide. The blue (A) and the red (B) spectra were taken 8.5 h and 16 h after the addition of enzyme, respectively. All spectra were set to the same contour level.

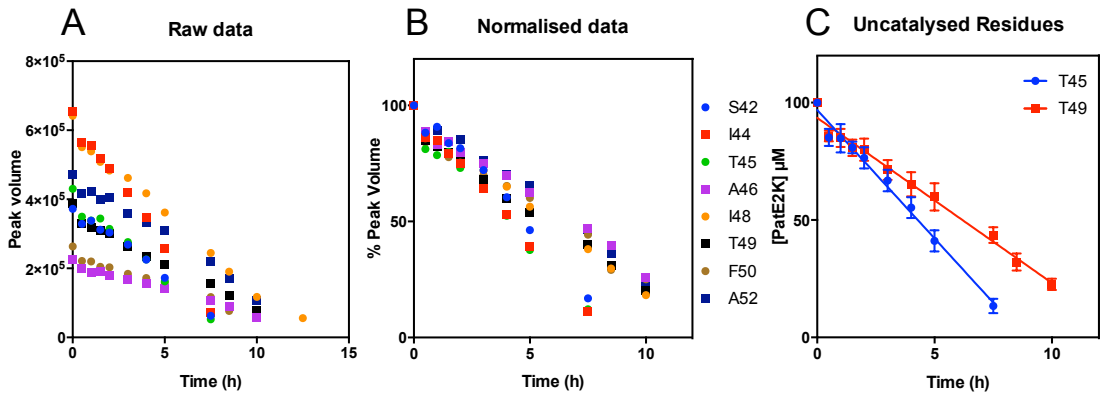

Figure 3.12: Rate of PatE2K-2het-MicD Reaction. (A) Intensities of cross peaks S42, K43, I44, T45, A46, I48, T49, F50, and A52 (original position) over the course of reaction. (B) Normalisation of data in (A) using maximum intensity of each dataset as 100% and 0 as 0%. (C) Normalised intensities of S42, I44 and T45 are averaged to monitor T45 heterocyclization; those of T49, F50 and A52 are averaged to monitor T49 reaction. Error bars shows standard errors of mean. Points were fitted to linear regression in Graphpad Prism.

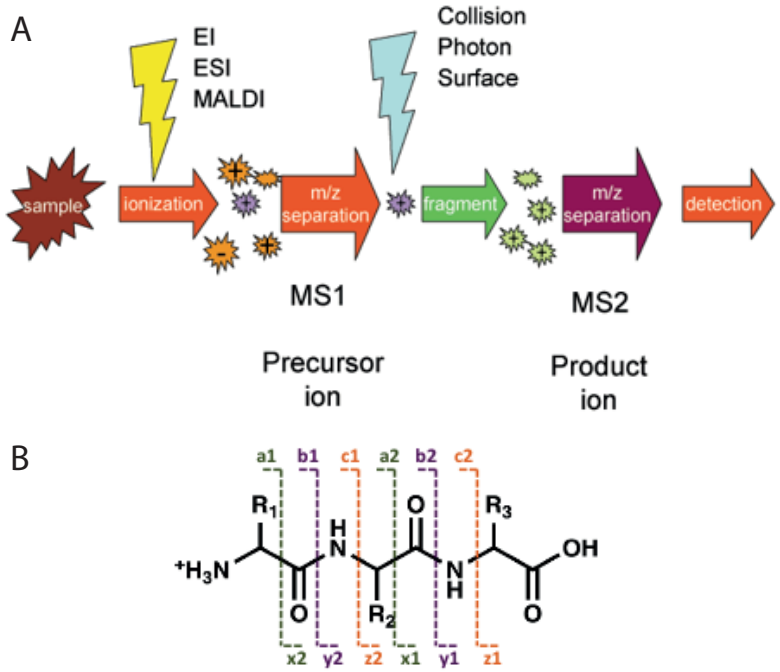

Figure 3.13: Simple overview of the principles underlying tandem mass spectrometry. (A) Ions are separated by mass and specific masses can be selected for fragmentation and secondary analysis. (B) Ion series resulting from fragmentation. b and y ions are used for analyses in this chapter. Figures taken from nationalmaglab.org.

Table 3.1: Possible heterocyclized derivatives of ITACITFCAYDG(EK) and their nomenclature. C^{het} denotes a thiazoline formed from a cysteine and its preceding amide bond, whereas T^{het} denotes a methyloxazoline formed from a threonine and its preceding amide bond.

Denotation	Sequence	Location of heterocycle(s)
I	ITACITFCAYDG(EK)	N/A
II	ITACITFC^{het}AYDG(EK)	C8
III	ITAC^{het}ITFCAYDG(EK)	C4
IV	ITAC^{het}ITFC^{het}AYDG(EK)	C4, C8
V	IT^{het}AC^{het}ITFC^{het}AYDG(EK)	T2, C4, C8
VI	ITAC^{het}IT^{het}FC^{het}AYDG(EK)	C4, T6, C8
VII	IT^{het}AC^{het}IT^{het}FC^{het}AYDG(EK)	T2, C4, T6, C8

Table 3.2: Calculated mass of b ions generated from peptide species listed in Table 3.1.

Reisude	ion	I	II	III	IV	V	VI	VII
I	b1	114.09	114.09	114.09	114.09	N/A	114.09	N/A
T	b2	215.14	215.14	215.14	215.14	197.13	215.14	197.13
A	b3	286.18	286.18	N/A	N/A	N/A	N/A	N/A
C	b4	389.19	389.19	371.18	371.18	353.16	371.18	353.16
I	b5	502.27	502.27	484.26	484.26	466.24	N/A	N/A
T	b6	603.32	603.32	585.31	585.31	567.29	567.29	549.28
F	b7	750.39	N/A	732.38	N/A	N/A	N/A	N/A
C	b8	853.40	835.39	835.39	817.37	799.36	799.36	781.34
A	b9	924.43	906.42	906.42	888.40	870.39	870.39	852.37
Y	b10	1087.50	1069.49	1069.49	1051.47	1033.46	1033.46	1015.44
D	b11	1202.52	1184.51	1184.51	1166.49	1148.48	1148.48	1130.46
G	b12	1259.54	1241.53	1241.53	1223.51	1205.50	1205.50	1187.48
E	b13	1388.59	1370.58	1370.58	1352.56	1334.55	1334.55	1316.53
K	b14	1516.68	1498.67	1498.67	1480.65	1462.64	1462.64	1444.62

Table 3.3: Calculated mass of y ions generated from ITACITFCAYDGEK-derived peptide.

Reisude	ion	I	II	III	IV	V	VI	VII
K	y1	147.11	147.11	147.11	147.11	147.11	147.11	147.11
E	y2	276.16	276.16	276.16	276.16	276.16	276.16	276.16
G	y3	333.18	333.18	333.18	333.18	333.18	333.18	333.18
D	y4	448.20	448.20	448.20	448.20	448.20	448.20	448.20
Y	y5	611.27	611.27	611.27	611.27	611.27	611.27	611.27
A	y6	682.30	682.30	682.30	682.30	682.30	682.30	682.30
C	y7	785.31	N/A	785.31	N/A	N/A	N/A	N/A
F	y8	932.38	914.37	932.38	914.37	914.37	914.37	914.37
T	y9	1033.43	1015.42	1033.43	1015.42	1015.42	N/A	N/A
I	y10	1146.51	1128.50	1146.51	1128.50	1128.50	1110.48	1110.48
C	y11	1249.52	1231.51	N/A	N/A	N/A	N/A	N/A
A	y12	1320.56	1302.55	1302.55	1284.53	1284.53	1266.52	1266.52
T	y13	1421.61	1403.60	1403.60	1385.58	N/A	1367.57	N/A
I	y14	1534.69	1516.68	1516.68	1498.66	1480.65	1480.65	1462.63

time (Figure 3.14A). Fragmentation pattern from the three peaks were similar, with the later peak (12-12.5 min) giving the highest ion intensities (Figure 3.14B). MSMS spectra of the other two peaks are found in Appendix B (Figure S3). IT^{het}AC^{het}ITFC^{het}AYDGEK and ITAC^{het}IT^{het}FC^{het}AYDGEK share many common fragments, and hence only ions b_1 to b_5 (Table 3.2) and y_9 to y_{13} (Table 3.3) are used to differentiate the two. These masses are plotted as stars (b ions) and circles (y ions) at given x values in order to signify peaks that correspond to each intermediate (Figure 3.14B). The strongest peaks are those that match y_9, b_4 and b_5 of peptide V (Table 3.1), suggesting that T2 (T45) has been heterocyclized, whereas T6 (T49) remains unmodified. Particularly, y_9 would have disappeared if T4 was cyclised. There is a small, but noticeable peak (around 2.5% intensity) with the mass of y_{13} of peptide VI, which could be noise, or an indication that a small amount of triply heterocyclized peptide has been modified on T49. However, it can be concluded that in the majority of cases, T45 is transformed before T49, giving an overall reaction order of C51-C47-T45-T49.

3.4.6 Reaction Order of MicD is Partially Leader-Dependent

Having established the reaction order when all four heterocycles are formed, we then investigated whether this is leader-dependent or intrinsic to the heterocyclase enzyme. First, an alkylation experiment was carried out to determine if chemoselectivity for cysteines still predominate. ITACITFCAYD was reacted with MicD fusion (MicD R16 5GA) and analysed by MALDI MS, before and after 5 mM IAA was added to the sample (Figure 3.15). After 1 h of reaction, the sample contained a mixture of singly, doubly and triply heterocyclized peptide, as well as the starting material (Figure 3.15A). Alkylation of this sample led to a +57 Da mass shift of the 1het species and a +114 Da mass shift of the unreacted peptide (Figure 3.15B), suggesting that like full length PatE2K, leaderless peptides are catalysed on the two cysteine residues first.

Next, the reaction order between the two cysteines was studied. ITACITFCAYDG was reacted with MicD fusion for 1 h to obtain singly heterocyclized ITACITFCAYDG-1het ($[M+2H]^{2+}$=630.3 Da). Meanwhile, PatE3KK (contains full leader sequence) was reacted with native MicD, followed by tryptic digest to yield ITACITFCAYDGEK-1het ($[M+2H]^{2+}$=758.8 Da). The former and latter peptide will be aliased as leaderless and 'full-length', respectively, to reflect their state when heterocyclization took place, although when subjected to MSMS they were both devoid of the leader peptide. Both species were selected and fragmented so as to determine the site of heterocyclization (Figure 3.16). As it has been proven that the C8/C51 was catalysed first when native MicD reacts with full-length PatE (Figure 3.7), the full-length peptide was used as a control, against which the leaderless peptide was analysed. It needs to be noted that the two peptides are not exactly the same, so there are natural variations between the ions resulting

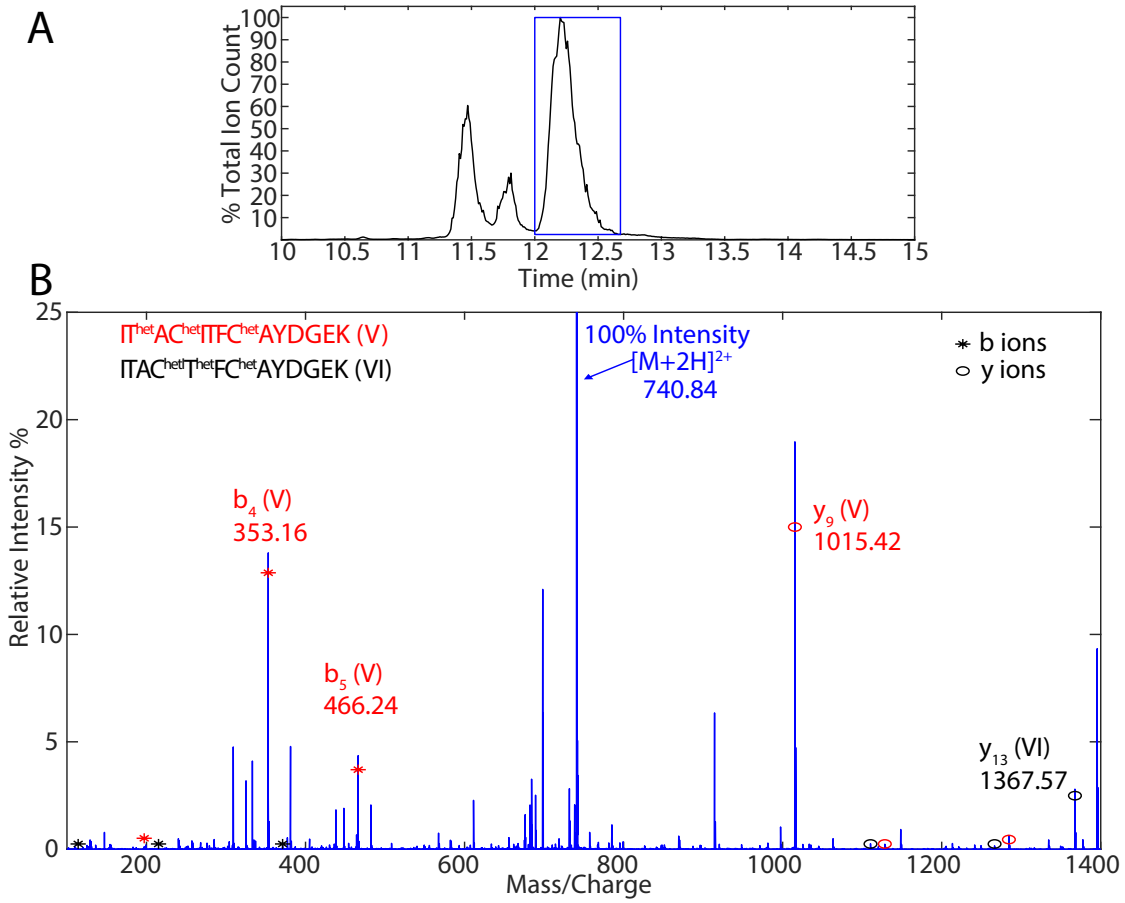

Figure 3.14: Fragmentation of Triply Heterocyclized ITACITFCAYDGEK. PatE3KK was reacted with MicD and cleaved by trypsin to release the peptide ITACITFCAYDGEK, which was subjected to LC-MSMS. (A) shows total ion count of the triply heterocyclized peptide (740.84 Da); blue box highlights the peak that was used to generate fragmentation data. (B) shows daughter ions generated from the 740.84 precursor that eluted between 12-12.5 min.

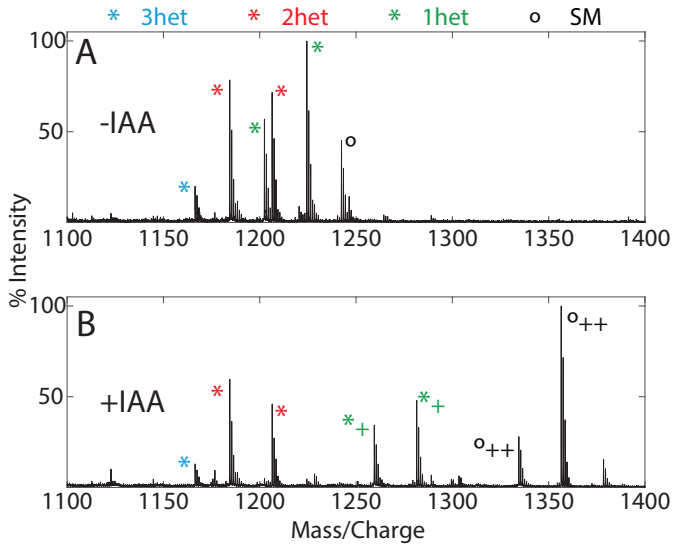

Figure 3.15: IAA Alkylation of Leaderless Peptide Reacted with MicD. ITACITFCAYD (100 µM) was reacted with MicD R16 5GA (5 µM) for 1 h at 25 °C. (A) MALDI mass spectrum before addition of IAA. (B) MS after the sample was reacted with 5 mM IAA for 30 min. $[M+H]^+$ species are labelled on their left and $[M+Na]^+$ species are labelled on their right.

from their fragmentation. B ions (b_1 to b_{12}) generated from ITACITFCAYDG- and ITACITFCAYDGEK-derived peptides share the same masses (Table 3.2), whereas y ions differ. Calculated masses of y ions originating from the ITACITFCAYDG series are listed in Table 3.4.

Mapping ions that differentiate peptide II (C8/C51-heterocyclic) from peptide III (C4/C47-heterocyclized), namely b_3 to b_7 and y_5 to y_9 for the leaderless peptide, onto the MSMS spectrum showed that b_4 (II), b_5 (II), y_6 (II) and y_7 (II) yielded the highest peaks (Figure 3.16A). Other ions, corresponding to both peptide II and III, were observed at lower intensities. Similarly, ions b_3 to b_7 and y_9 to y_{13} were annotated on the spectrum of the 'full-length' peptide (Figure 3.16B); the strongest signals were again attributed to those generated from peptide II (b_4 (II), b_5 (II), y_9 (II)). Even though this sample was expected to be uniformly C51/C8 heterocyclized, fragmentation of it still yielded ions that matched masses of b and y ions from peptide III. These could be fragments resulting from tryptic digest of the enzyme that coincided with the masses of interest, or they could have arisen due to hydrolysis of the C8-derived thiazoline on an originally doubly heterocyclized peptide during the LC stage, as the solvent contained 0.05% TFA. Regardless, the influences should be similar on the MicD + PatE3KK (Figure 3.16B), MicD fusion + ITACITFCAYDG (Figure 3.16B) samples. As fragmentation of the leaderless peptide was overall more extensive than that of the 'full-length' one, direct comparison of ion intensities was not drawn. However, a pairwise comparison of b ions was made focusing on the ratio of II:III (Figure 3.16C), where the magnitude of the MSMS spectrum of the

Table 3.4: Calculated mass of y ions generated from ITACITFCAYDG-derived peptide species, following nomenclature given in Table 3.1.

Reisude	ion	I	II	III	IV	V	VI	VII
G	y1	76.04	76.04	76.04	76.04	76.04	76.04	76.04
D	y2	191.07	191.07	191.07	191.07	191.07	191.07	191.07
Y	y3	354.13	354.13	354.13	354.13	354.13	354.13	354.13
A	y4	425.17	425.17	425.17	425.17	425.17	425.17	425.17
C	y5	528.18	N/A	528.18	N/A	N/A	N/A	N/A
F	y6	675.24	657.23	675.24	657.23	657.23	657.23	657.23
T	y7	776.29	758.28	776.29	758.28	758.28	N/A	N/A
I	y8	889.38	871.37	889.38	871.37	871.37	853.35	853.35
C	y9	992.39	974.38	N/A	N/A	N/A	N/A	N/A
A	y10	1063.42	1045.41	1045.41	1027.39	1027.39	1009.38	1009.38
T	y11	1164.47	1146.46	1146.46	1128.44	N/A	1110.43	N/A
I	y12	1277.55	1259.54	1259.54	1241.52	1223.51	1223.51	1205.49

'full-length' peptide was amplified five times for illustrative purposes. The II:III ratios were consistent between the two, indicating that N-to-C order of thiazoline formation was retained in the absence of an *in-cis* leader peptide.

The order of threonine (T2/T45 and T6/T49) heterocyclization when the leader is not within the precursor peptide was then investigated. ITACITFCAYDG was reacted with MicD fusion (Q21 5GA) for 16 h, and after LC-MS the mass that matched the triply heterocyclized intermediate was subjected to MSMS. It was reasoned that the order of methyloxazoline formation could be determined based ions b_2, b_4, y_8 and y_{10} (Table 3.2, Table 3.4), which have different masses in peptide V and VI (Table 3.1), as well as b_1, b_5, y_7 and y_{11}, which are only present in one of the triply heterocyclized peptides. Four peaks were observed during LC (Figure 3.17A) and analysed individually. Raw spectra of them are found in Appendix B. Although intensities of all peaks of interest were low (below 5%, with the intensity of the unfragmented precursor as 100%), trends were observed when they were examined in detail (Figure 3.17B). The earlier eluting peaks (peak 1 and 2, Figure 3.17A) exhibited similar fragmentation patterns between the two (Figure 3.17B), and so did the peaks with longer retention times (peak 3 and 4, Figure 3.17A) to a less extent. In LC-MS peaks 1 (blue) and 2 (green), neither b_2 ions were detectable, while a high b_4(V):b_4(VI) ratio (about 5:1 in terms of peak height) pointed to T2 heterocyclization in these species. This was however contradicted by the reverse V:VI ratio (about 1:3) of y_8. In LC-MS peaks 3 (red) and 4 (orange), fragments were biased towards VI, as shown by the b_2, y_{10} and y_{11} ions. The MSMS peak for b_4(V) was stronger than that for b_4(VI), but the ratio is much lower than that observed for the early eluting species (green and blue), again indicating that peptide V was less abundant at least in the later eluting species (red and orange). Like those from peaks 1 and 2, y_8 originated from peaks 3 and 4 showed a reverse pattern, with y_8(V) being stronger than y_8(VI). Attempts were made to obtain clearer MSMS spectra, such as varying collision voltage, but the effect was limited. Based

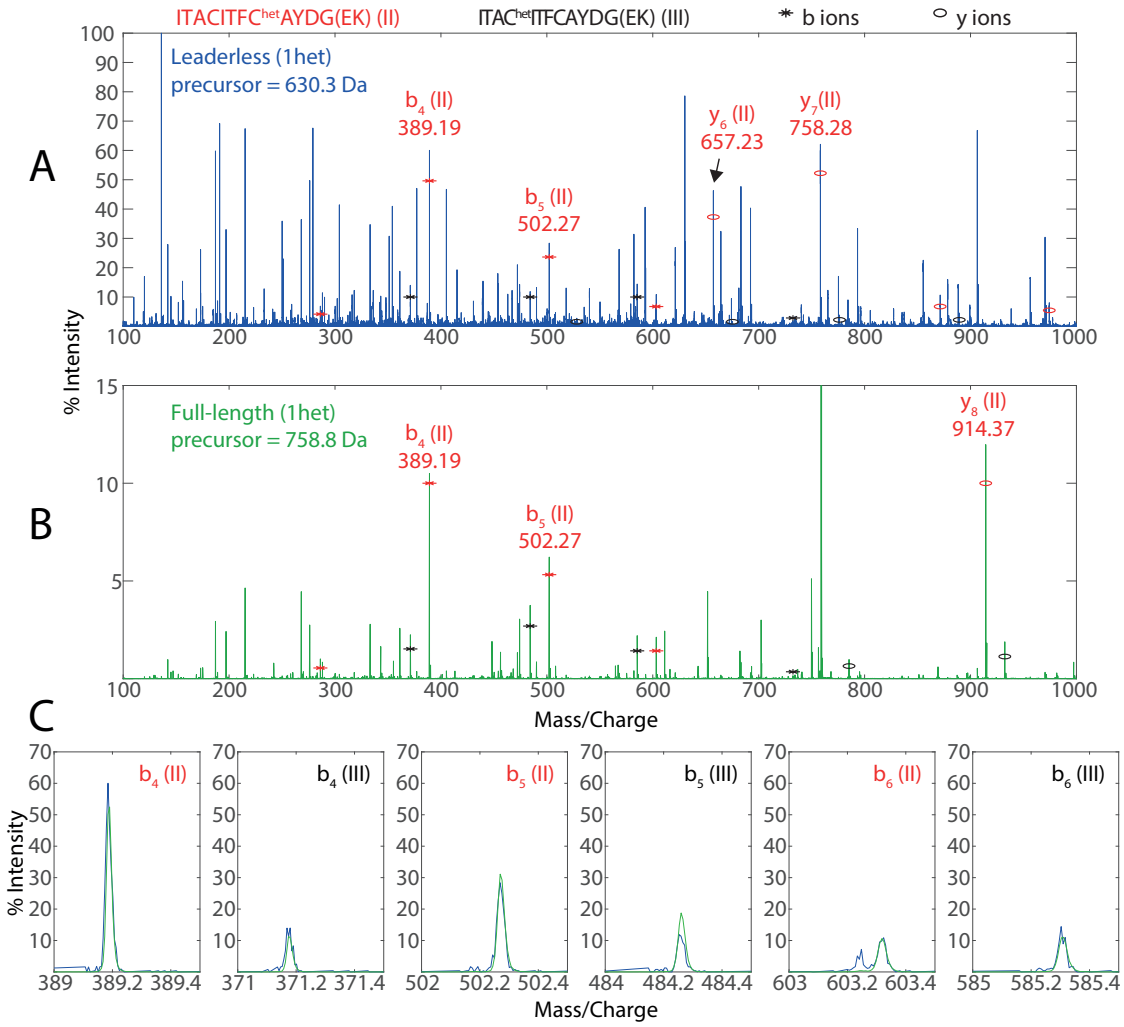

Figure 3.16: MSMS of Singly Heterocyclized ITACITFCAYDG(EK). (A) MSMS spectrum of singly heterocyclized ITACITFCAYDG processed by MicD fusion. (B) MSMS spectrum of singly heterocyclized ITACITFCAYDGEK that has been reacted with native MicD. Both annotated with b ions (stars) and y ions (circles) that differentiate peptide II (red) and III (black). (C) Pairwise comparison of II:III ratios of native (green) and fusion (blue) MicD processed peptides. As the green spectrum is overall weaker its magnitude is amplified five times, hence the intensity should not be compared between the two different spectra, but only between the same spectrum in different mass regions.

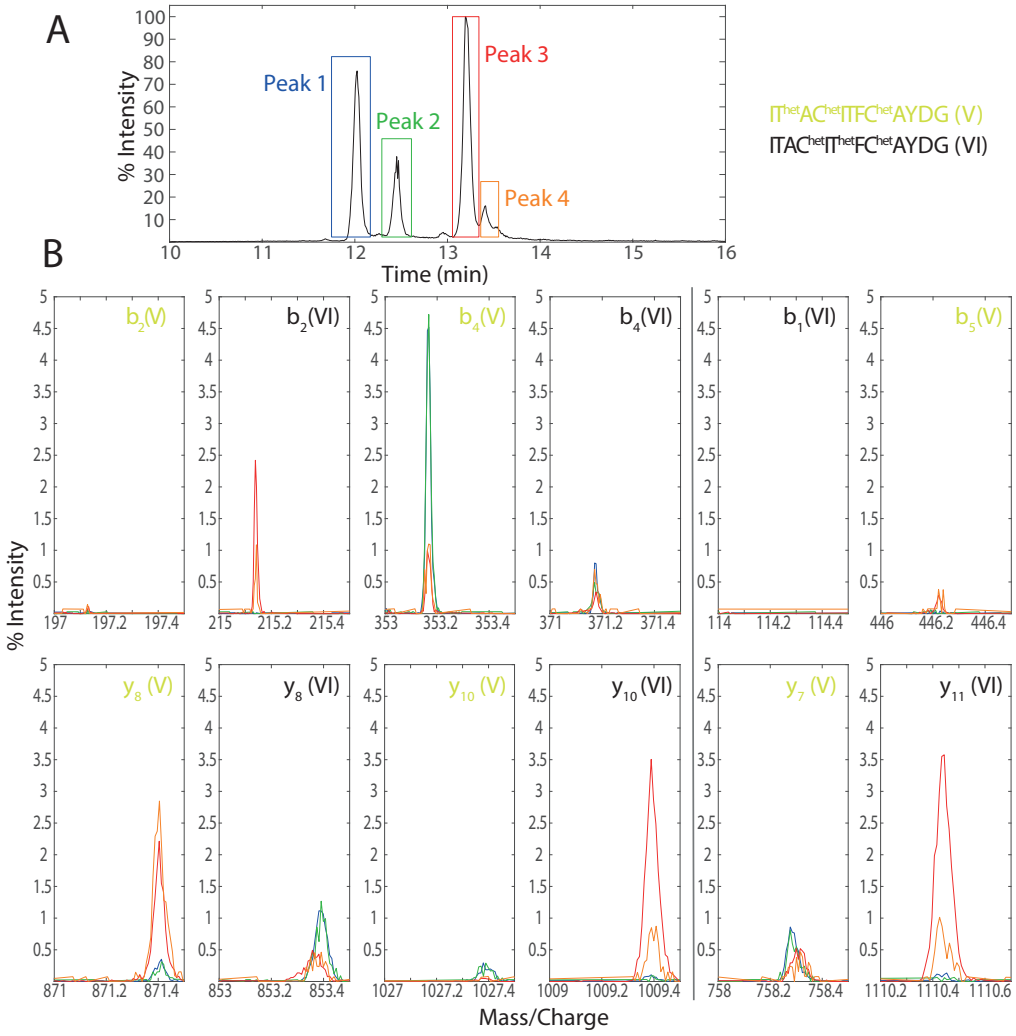

Figure 3.17: MSMS of triply heterocyclized ITACITFCAYDG. (A) TIC of ions with the mass 612.3 Da. Boxed peaks gave MSMS spectra of corresponding colours in (B). (B) MSMS spectra of different peaks zoomed to mass regions that differentiate peptide V from peptide VI. Peak 1, blue; peak 2, green; peak 3, red; peak4, orange.

on the available results it was concluded that the later eluting peaks contained, whether or not exclusively, triply heterocyclized peptides with T6-derived methyloxazolines. Hence, the reaction order with regard to threonines appears lost when the leader is not attached to the core peptide.

3.4.7 LynD Heterocyclizes Both Cysteines on the Linear Substrate

As the partial leader-dependence of reaction order observed in MicD-catalysed heterocyclization was unexpected, another homologue, LynD, was examined with regard to the sequence of thiazoline formations from C4 and C8. Interestingly, native LynD did not exhibit preference for either cysteine residue, as shown by MSMS analysis of the singly

heterocyclized intermediate of PatE3KK which has been reacted with LynD for 1 h and then trypsinated (Figure 3.19A, Figure 3.18A). A good signal to noise ratio was seen in the 700-1500 m/z region of this experiment, with prominent and distinct peaks for y_7 (III), y_8 (II), y_8 (III), y_9 (II) and y_9 (III), y_{10} (II) and y_{11} (II). Particularly, y_8 and y_9 ions from peptide II and III showed very similar intensities, suggesting that ITACITFCAYGEK-1het consisted of both species. This observation was unanticipated both TruD [65] and MicD operate in an defined C-to-N order . Unsurprisingly though, loss of *in-cis* leader did not result in gain of reaction order, demonstrated by LynD fusion not preferring either C4 or C8 on ITACITFCAYDG (Figure 3.18B). Again, the more intense peaks matched ions from both peptide II and III, in contrast to the spectra of MicD-reacted peptides which showed biases towards II (Figure 3.16A and B).

When the reaction was allowed for another 2 h, the reaction progressed and more heterocycles were installed (Figure 3.19B). MS/MS of the 1het species from this sample gave a highly similar fragmentation pattern (Figure 3.19C), meaning that neither ITAC^{het}ITFCAYDGEK nor ITACITFC^{het}AYDGEK was preferentially used in the second heterocyclization event, further confirming the random order of LynD-catalysed reactions.

3.4.8 Leader Peptide Dominates Binding

The disorder of LynD-catalysed reactions further showcased the great flexibility of the cyanobactin heteorcyclase with regard to the sequence of the core peptide. To understand the relative importance of core sequence and leader sequence to binding, steady-state kinetics of native and fusion heterocyclases-catalysed reactions were compared using a commercially available assay that monitors the rate of phosphate/pyrophosphate release (described in chapter 2, section 2.3.11). As the assay was carried out on much shorter time scale (10 min) than the NMR and mass spec experiments, it was thought to primarily report on the first heterocyclization. LynD and MicD were reacted with PatE2K, whereas LynD fusion and MicD fusion (R16 5GA) were reacted with ITACITFCAYDGE and ITACITFCAYD, respectively. It was postulated that if the core dominates binding initially, similar kinetic parameters would be observed when the enzyme is reacted with full-length PatE2K and its leaderless counterpart. Indeed, when the observed data fitted to the quadratic equation (Eq 2.2), k_{cat} values were consistent in the native and fusion enzymes-catalysed reactions (Figure 3.20A and B), although it differed between the enzymes with LynD and LynD fusion being faster than MicD and MicD fusion. K_m on the other hand varied significantly when the leader is present or absent *in cis*, with native MicD and LynD exhibiting a nanomolar range K_m for PatE2K, whereas the fusion enzymes' K_m for leaderless substrates were nearly three orders of magnitude higher (Table 3.5). It was therefore likely that enzyme-substrate binding is dominated by the leader peptide from the beginning. Admittedly, the sequence difference between PatE2K and

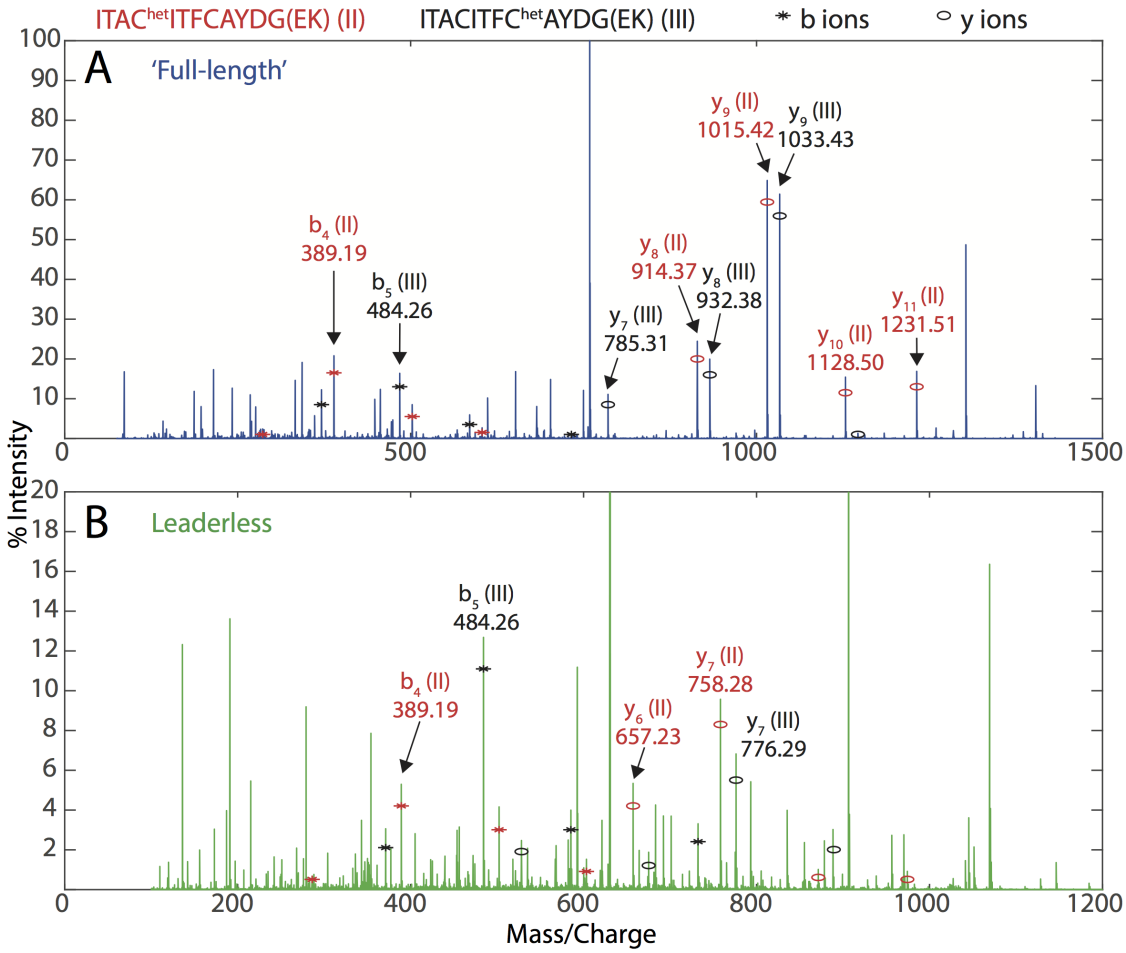

Figure 3.18: (A) ESI-MS/MS of trysinated PatE3KK which has been heteroccyclized once by LynD. (B) MSMS of ITACITFCAYDG-1het after reaction with LynD fusion.

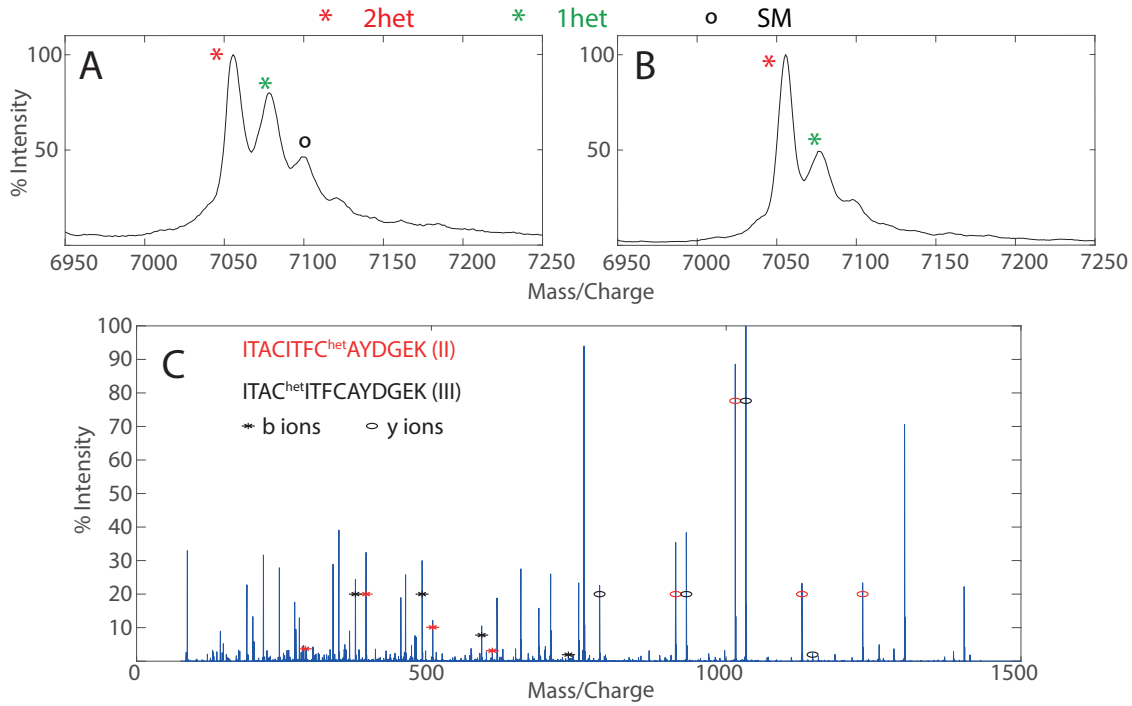

Figure 3.19: MALDI-MS of PatE3KK Catalysed by LynD Native. (A)(B) PatE3KK (unmodified MW = 7093 Da) was reacted with native LynD before 1 h (A) and 3 h (B) and subjected to MALDI-MS. (C) The 3 h sample was digested with trypsin and subjected to and ESI-MS/MS. Trace showing fragmentation of the ion species of 758.85 Da. SM, starting material; 1het, single heterocyclized; 2het, doubly heterocyclized.

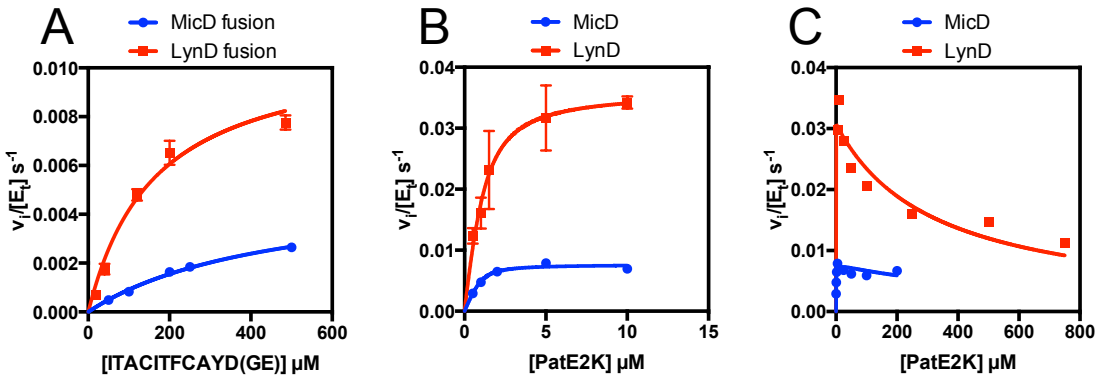

Figure 3.20: Steady-state Kinetics of Native and Fusion Heterocyclases. (A) LynD fusion (5 µM) was reacted with ITACITFCAYDGE whilst MicD fusion (Q21 5GA, 5 µM) was reacted with ITACITFCAYD. (B)Native MicD (1 µM) and LynD (0.4 µM) were reacted with 0.5 - 10 µM PatE2K. (C) Native MicD (1 µM) and LynD (2 µM) were reacted with higher concentrations of PatE2K. Data from (A) and (B) were fitted to Eq 2.2 (page 40) in Graphpad Prism; data from (C) were fitted to Eq 3.1.

ITACITFCAYD(GE) are large and other elements from PatE2K (apart from the leader) may also play a role in affinity, but this is unlikely as previous biophysical experiments have not demonstrated binding between other regions of PatE and TruD or LynD [64, 65]. It was noted that higher concentration of PatE2K appeared to have an adverse effect on the initial rate of reaction (Figure 3.20C), possibly due to substrate inhibition - preliminary data obtained from reacting LynD was high concentrations PatE2K (not in replicates) appeared to fit this model (eq 3.1), but further experiments are needed for confirmation. Values of k_{cat} and K_m listed in Table 3.5 were derived from the ascending phase (Figure 3.20B) only.

Table 3.5: Steady-state kinetics of MicD and LynD with full-length and leaderless substrates.The units of k_{cat} and K_m are in s^{-1} and µM, respectively, while k_{cat}/K_m is shown in s^{-1} μM^{-1}.

Enzyme	substrate	k_{cat}	K_m	k_{cat}/K_m
MicD	PatE2K	0.007 ± 0.0003	0.20 ± 0.07	0.035
LynD	PatE2K	0.036 ± 0.0005	0.53 ± 0.33	0.25
MicD fusion	ITACITFCAYD	0.005 ± 0.0005	459 ± 74	$1.1x10^{-5}$
LynD fusion	ITACITFCAYDGE	0.011 ± 0.0014	160 ± 48	$6.9x10^{-5}$

Another interesting observation, which corroborates the importance of the leader peptide in binding, was the inability of MicD fusion to further modify doubly heterocyclized leaderless peptides. Along with the MSMS experiments described in section 3.4.6, NMR spectra were collected to monitor the changes of ^{15}N-PatE2K-2het-tryp (ITACITFCAYDGELEHHHHHH) after the addition of MicD fusion, in order to determine the reaction order between threonines. No changes signifying cyclodehydration was observed after 16 h incubation at 20 °C (Figure 3.21). This was surprising, as in the MSMS

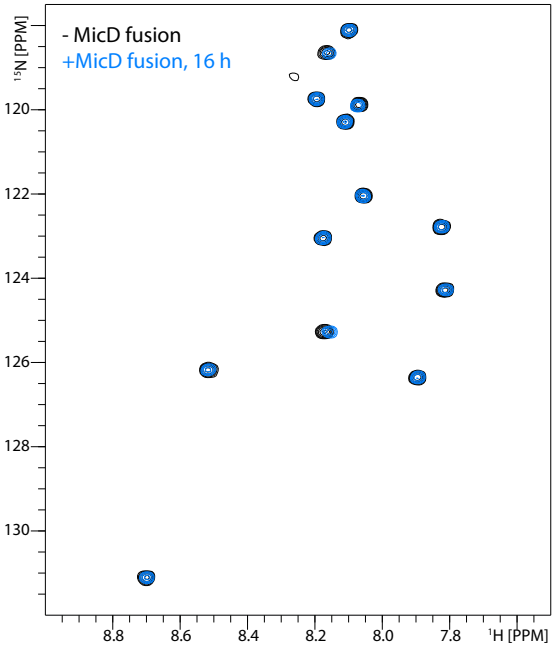

Figure 3.21: ^{1}H,^{15}N-HSQC spectra of PatE2K-2het-tryp before (black) and after 16 h reaction (blue) with MicD R16 5GA. This peptide was also used to study oxidation, as described in chapter 4, where sequential assignment is shown. Here, no difference was observed before and after the reaction.

experiments triply heterocyclized peptides were made from linear ITACITFCAYDG (Figure 3.17A). It was thought that the pH 7.4 condition used in the NMR experiment might hindered the reaction (see chapter 2), and parallel reactions were set up with ^{15}N-PatE2K-2het-tryp and MicD R16, in either 50 mM HEPES pH 7.4 or 50 mM Tris pH 8.0 (Figure 3.22A and B). Neither reaction showed further heterocyclization when subjected to MALDI-MS. To ensure the enzyme was active, a control with and the fully linear version of PatE2K-tryp; in this reaction a significant 3het species was observed (Figure 3.22C).

3.5 Discussion

The current study has expanded the knowledge on the order of the heterocyclases from the cyanobactin family. Whilst it was originally thought these enzymes would all exhibit a C-to-N order of reaction, the results surprisingly showed that individual enzymes differ in their preferences with regard to heterocyclizable residues, despite their relatedness (Figure 1.10). Being a catalyst of both thiazolines and (methyl)oxazolines, MicD is chemoselective for thiols over alcohols. It proceeds from the C- to the N- terminus like TruD, but then takes a U-turn after installing the first methyloxazoline. LynD, evolutionarily more distant to TruD and MicD (Figure 1.11), interestingly showed a natural disorderedness (Figure 3.18). Determining the reaction orders of other heterocyclases such as PatD and OscD may

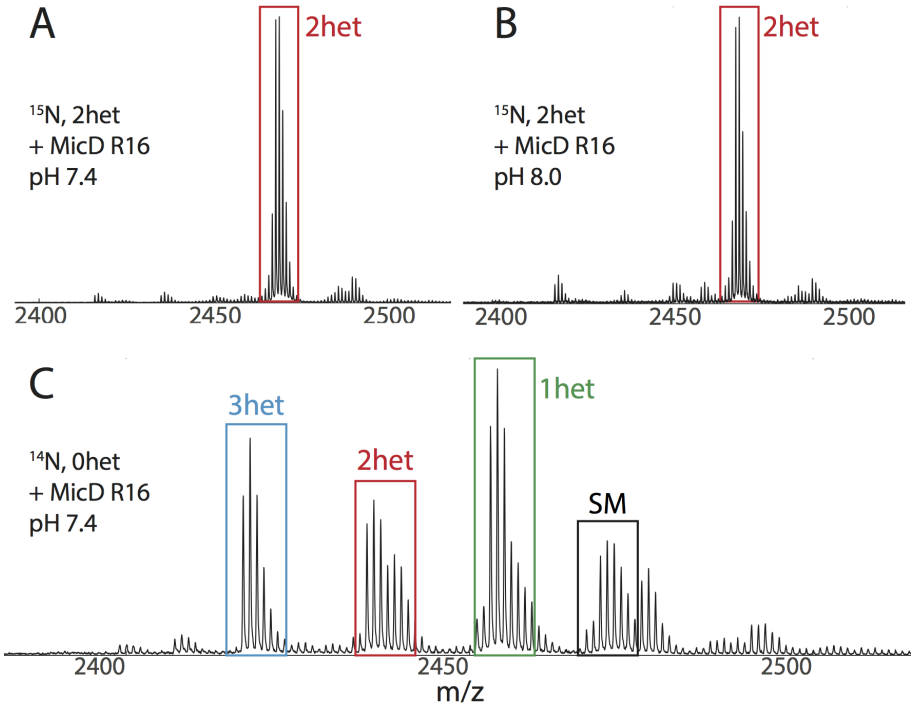

Figure 3.22: Trypsin Cleaved PatE2K Reacted With MicD. (A)(B) ^{15}N-PatE2K-2het-tryp was reacted with MicD R16 5GA in a reaction buffer containing 50 mM HEPES pH 7.4 and 50 mM Tris pH 8.0, respectively. (C) Unlabelled PatE2K-tryp (no heterocycle) was reacted with MicD R16.

provide a more convincing link between genotypes and phenotypes. The variation between homologous enzymes may also apply to other RiPP systems. For example, there was a discrepancy in the order of dehydrogenation when the oxidase BcerB was reacted with pre-heterocyclized peptides, and when the BalhBCD complex was used to heterocyclize and oxidise the precursor (see section 3.2) [113]. The authors argued that the BalhCD (heterocyclase) reaction order dominates when the complex is formed, as BalhCD and BalhBCD showed the same overall C-to-N directionality. However, since the oxidation reaction was carried out with BcerB, intrinsic differences between this enzyme and BalhB may also have caused the change in order.

The leader peptide appeared to be involved in determining the reaction order, but is not the sole factor. The heterocyclization order of MicD was partially abolished - changing from C8-C4-T2-T6 to C8-C4-T2/T6-T6/2, when the leader peptide was detached from the core. This observation makes it less likely that heterocycles flanking T6 were restricting its accessibility; rather, orientation of the core in the active site may differ when the leader is and is not docking it to the protein, with the leaderless peptide having a larger degree of freedom. A similar reset of order was seen in the lantibiotic synthetase LctM (see section 3.2) [88], although LctM is not only partially ordered to begin with. Reaction order may arise from reduced or increased substrate affinity for

each subsequent intermediate, associated with changes in chemistry and conformation. Haloduracin (lantibiotic) biosynthetic enzyme HalM2, which is highly substrate specific, exhibits increased rates of thioether ring production as the precursor peptide matures [162]. In contrast, ProcM, a permissive lantibiotic synthetase, cyclises the peptide at a reduced rate after the first reaction [162]. Like in ProcM, relaxed selectivity of cyanobactin heterocyclases is coupled to the decreasing rate of reaction as the number of heterocycles grow [64, 65]. This trade-off between catalytic and efficiency substrate tolerability was also seen in other RiPP PTMEs; for example, the promiscuous cyanobactin macrocyclase PatG is drastically slower than its more substrate-specific counterpart from plants, PCY1 [45, 91, 163]. Rather than maintaining reaction order, the leader peptide plays a more crucial role promoting substrate binding. The K_m of MicD/LynD for full-length PatE is several orders of magnitude lower than that of MicD fusion/LynD fusion for leaderless peptide with the same core sequence (Figure 3.20). At first this appears comparable to the pharmacological concepts of avidity and target residence time, where a a bi/multivalent ligand such as an antibody binds to two sites on the same target protein [164]. The binding of the first pharmacophore increases the apparent affinity of the second (avidity) due to forced spatial proximity, as well as enhancing the likelihood of the conjoined ligand rebinding to the protein once it has dissociated (longer target residence time), for the same reason. The leader and core peptides are considerably farther apart than epitopes on an immunoglobin, but leader binding still limits the diffusion of core peptide into a smaller space, thereby increasing the its local concentration. The latter effect, however, would not benefit the first heterocyclization, or the subsequent one as of TruD [65] and MicD 3.7, both of which are distributive, meaning that after modification of C8 on one peptide they go onto the next peptide to carry out the same reaction, rather than switching to another residue on the same peptide. Hence, the leader peptide dissociates alongside the core peptide after the first reaction. Avidity on the other hand could assist in the binding of the linear and heterocyclized core peptides, as leader biding brings them into proximity of the active site. Nevertheless, even when it is not adjacent to the core, leader peptide still activates the heterocyclase, demonstrated by the elevated reaction rate of fusion enzymes (chapter 2) and *trans*-activated enzymes [64] towards leaderless peptides compared to native enzymes. This might be viewed as allosteric activation, by increasing the enzyme's affinity for the core and/or its catalytic efficiency as the crystal structure of LynD shows the leader making specific contacts with the protein, which is conformationally different from its apo form (as of TruD) [64, 65]. The schematic below shows the hypothetical kinetic pathways involved in allosteric activation (Figure 3.23). For simplicity, ATP is assumed to bind to the enzyme first, and substrate inhibition is not taken into account. Fusion enzymes can be seen as the activator-bound form ENL, under the premises that the leader is constantly bound to the enzyme in the same way as when is it part of PatE - in reality, the efficiency of leader-binding in a fusion enzyme depends on the exact architecture of that protein, but after testing various fusion constructs similar kinetic parameters were obtained (chapter 1), all of which more or less represent K_{m2} and k_{cat2}. Comparing k_{cat2}

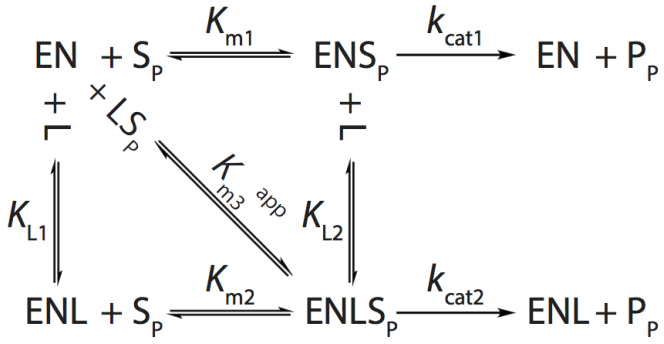

Figure 3.23: Kinetic model for leader peptide induced allosteric activation. For simplicity ATP is assumed to bind first. EN, nucleotide-bound enzyme; S_p, core peptide; L, leader peptide; P_p, heterocyclized core peptide. K_{L1} and K_{L2}, dissociation constants for the leader peptide.

to k_{cat1} could reveal crucial information about the activation type - if these values are the same, only $ENLS_p$ is catalytically competent [165]. This is clearly not the case as native enzyme does process leaderless peptide, although the reaction rate is so slow that accurate velocity measurement using the phosphate assay is prevented due to noise from background ATP hydrolysis. K_{m1} and k_{cat1} have not been obtained. Kinetic details about leader-activation may be achieved by measuring the rate of reaction using *in trans* leader and core peptides, thereby allowing the variation of each component independently. Finally, the high velocity at apparently low concentrations of PatE (low K_{m3}^{app}) seen in the native enzyme assays (Figure 3.20), compared to leaderless peptides, have likely benefited from an avidity-type inflation of local concentration.

It remains a mystery why MicD (fusion) would not further process doubly heterocyclized leaderless peptides (Figure 3.21), as it does install methyloxazolines on peptides with 2 thiazolines when they have the leader attached (Figure 3.11), as well as on leaderless peptides that do not initially have any heterocycles (Figure 3.15). Another open question is whether the reduction in rate seen with higher amounts of PatE is a consequence of substrate inhibition. Further work needs to be carried out, covering a wide range of peptide concentrations (up to mM levels) to fully establish this. The decline in rate was not seen with the concentrations of leaderless peptides used (limited due to solubility issues), but may appear if concentration could be further increased.

Chapter 4

Effect of Leader Peptide On Azoline Oxidation

4.1 Summary

The cyanobactin oxidase contains a C-terminal FMN-containing, and an N-terminal domain, which contains two peptide clamps of unknown function(s). Due to the structural resemblance of the clamps in the N-terminal domain of oxidase to the peptide clamp in the cyanobactin heterocyclase LynD, which mediates leader peptide binding, it was hypothesised that these elements may perform the same function in the oxidase. Yet, employing NMR spectroscopy it was demonstrated that the enzyme ArtGox preferred peptide substrates that do not have a leader sequence; in addition, oxidation proceeds in a specific order that is not altered by the cleavage of leader peptide. The oxidase's preference for leaderless substrates suggests that oxidation takes place after proteolysis of the leader peptide during cyanobactin biosynthesis. These findings fit with other biochemical and biophysical data in establishing that the peptide clamps in the oxidase are vestigial and do not participate in leader binding.

4.2 Introduction

Azolines within cyanobactin precursors are frequently subjected to oxidation by an FMN-dependent protein, known as the oxidase or dehydrogenase (Figure 1.15). The general differences between oxidised and unoxidised heterocycles are discussed in section 1.5. The effect of azoline oxidation in cyanobactins is exemplified by lissoclinamide 5, which differs from lissoclinamide 4 only by the change of one thiazoline to thiazole, yet demonstrates a two orders of magnitude increase in cytotoxicity against bladder carcinoma cells compared

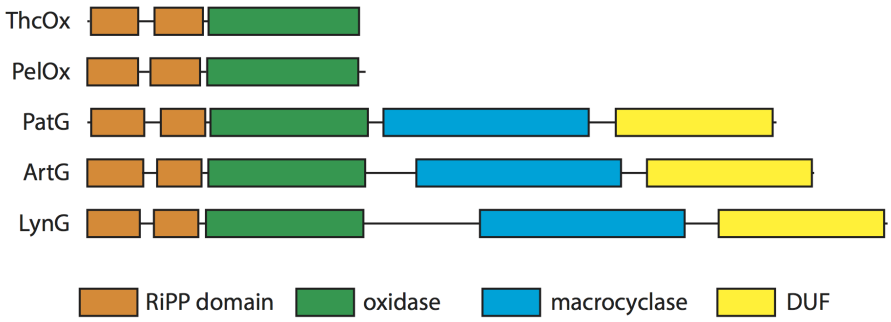

Figure 4.1: Schematic showing domains in cyanobactin oxidases. ThcOx (*Cyanothece* sp. *PCC 7425*), standalone oxidase. PelOx (*Nostoc* sp. *Peltigera membranacea cyanobiont 232*), putative standalone oxidase. PatG (*Prochloron didemni*), ArtG and LynG (*Lyngbya* sp. *PCC 8106*), fused oxidases. DUF, domain of unknown function. Domains were assigned by the HHpred server [167].

to the latter [166]. Cyanobactin oxidases, as opposed to their equivalents in other RiPP and NRP pathways, uniquely possess two peptide clamp motifs (also known as RiPP recognition elements, RRE, or RiPP domains) in addition to the FMN-containing oxidase domain. These two clamps appear to be prevalent in cyanobactin oxidases, regardless if they are standalone enzymes or fused to the G protein (Figure 4.1), but their function remains unknown. In the crystal structure of ThcOx [106], one of them is seen buried in the dimeric interface, while the other is exposed to solvent. Although it was hypothesised that the solvent-accessible clamp may engage in leader peptide binding in a similar fashion to the peptide clamp in LynD [64], protein-peptide interaction was not observed in the crystal structure of ThcOx. This chapter describes the biochemical and biophysical studies carried out to validate the potential effect of leader-binding on the activity of the oxidase. In order to achieve that aim, the reaction rate and order of ArtGox (*Arthrospira platensis*) reacted with full-length and leaderless substrates were determined, compared and contrasted.

The enzyme used, ArtGox (*Arthrospira platensis*), is a fused (to macrocyclase) cyanobactin oxidase from the arthrospiramides biosynthetic pathway [119]. Four natural products, arthrospiramide A (GLPGMPCF), arthrospiramides B (GLVLMPCP) and a sulfoxide variant of each (Figure 4.2), have been isolated to date, all of which contain two thiazoles. (Methyl)oxazole(in)es were not found in these compounds, but serine and threonine residues were found in several ArtE core sequences whose natural products have not been isolated, leaving the possibility that natural products of ArtG could contain oxazoles [119]. The ability of ArtGox to oxidise oxazolines was thus investigated.

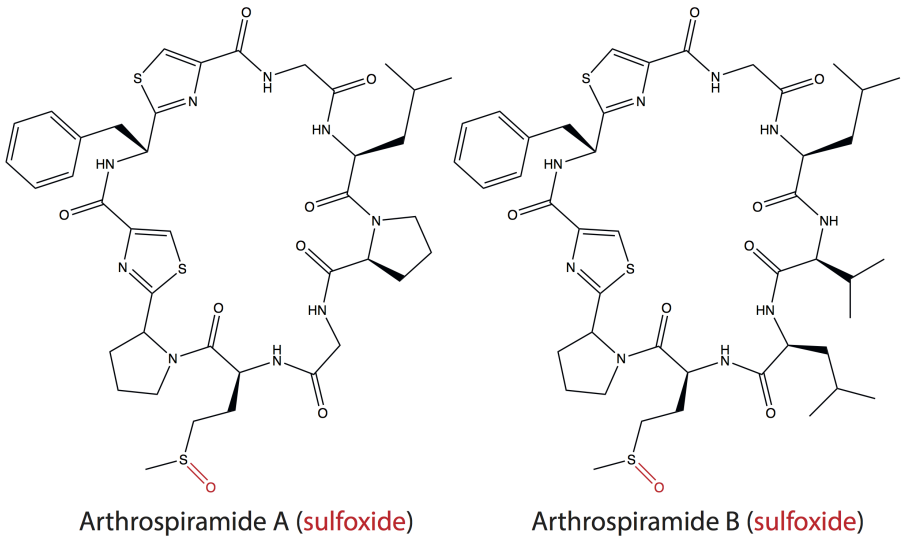

Figure 4.2: Structures of Arthrospiramides A, B and sulfoxide variants.

4.3 Materials and Methods

4.3.1 Overexpression and Purification of ArtGox

ArtGox, encoding the oxidase domain of ArtG, was cloned into pEHISTEVSUMO vector by Dr Andrew Bent, and expresses and a recombinant protein containing a N-terminal His-tag and SUMO-tag followed by a TEV cleavage site. The plasmid was transformed into *E. coli* BL21 (DE3) cells and recovered from a LB-kanamycin plate. A single colony was transferred to 100 mL LB-kanamycin medium and incubated at 37 °C for 16 h. 6 ml starter culture was then inoculated into 400 ml autoinduction medium [147] supplemented with 50mg/l kanamycin and 50 µM riboflavin. Cells were incubated at 20 °C, 250 rpm for 48 h and harvest by centrifuging at 8983g, 4 °C for 15 min. Pellets were resuspended in oxidase lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 3 mM BME and 50 µM flavin mononucleotide (FMN)) supplemented with complete EDTA-free protease inhibitor (one tablet per 50 ml of suspension) (Roche) and DNase (0.4 g per gram of cell pellet) (Sigma). The suspension was filtered through a sieve and passaged through a cell disruptor (Constant Systems Ltd), which was equilibrated with lysis buffer and chilled to 4 °C, at 30 kpsi. Cell lysate was cleared by centrifuging at 48384 g for 20 min at 4 °C, followed by filtration through a Millipore nitrocellulose membrane with a pore size of 0.45 µm (Sigma). Cleared lysate was applied to a Ni-NTA column (GE Healthcare), equilibrated with oxidase lysis buffer, and eluted with oxidase elution buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 250 mM imidazole, 3 mM BME and 50 µM FMN). The eluate was collected and concentrated where necessary, and subjected to size-exclusion chromatography on a 16/600 Superdex 200 pg column (GE Healthcare) in gel filtration buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM TCEP).

4.3.2 NMR Experiments with full-length E2K

^{15}N- and ^{13}C, ^{15}N-labelled PatE2K-2het were produced as described in section 3.3.3. Experiments were carried out on a Bruker AVANCE 500 MHz spectrometer equipped with a TXIz probe, controlled by the Bruker Topspin software. For oxidation monitoring, 0.2 M ^{15}N-labelled PatE2K-2het was reacted with 10 µM ArtGox in 10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM FMN and 5% D2O. HSQC spectra were also recorded at 10 °C for the starting material, and 12 h and 60 h after the start of reaction. Between these time points the sample was incubated at 20 °C. ^{1}H, ^{15}N HSQC spectra were acquired using a standard Bruker pulse sequence with Watergate water suppression with 8 transients at 2048×128 points with spectral resolutions of 9.8 Hz and 19.0 Hz in the direct (^{1}H) and indirect (^{15}N) dimensions, respectively. For sequential assignment, HNCACB spectra were acquired with 16 transients at 1024×50×128 points and spectral resolutions of 13.7 Hz, 44.6 Hz, and 157.1 Hz for the ^{1}H, ^{15}N and ^{13}C dimensions, respectively. CBCA(CO)NH spectra were acquired with 8 transients at 1024×50×114 points and spectral resolutions of 13.7 Hz, 44.6 Hz, and 176.4 Hz for the ^{1}H, ^{15}N and ^{13}C dimensions, respectively. Equivalent spectra were recorded for fully oxidized u-^{13}C, ^{15}N-PatE2k-2het (concentration 1.0 mM) and a sample where the enzymatic transformation was interrupted after 12 h by gel filtration on a 16/600 Superdex 75 pg column (GE Healthcare) in heterocyclization buffer supplemented with 5 mM FMN, thus allowing assignment of the reaction intermediate. Raw spectral data were processed with NMRPipe [158] and analysed with CCPN Analysis 2 [159]. Sequential assignment was carried out by Dr Uli Schwarz-Linek.

4.3.3 NMR Experiments with Leaderless E2K

^{15}N-labelled PatE2K-2het-tryp was prepared as described in 3.3.3. NMR experiments with the leaderless substrate were performed at 20 °C on a Bruker Ascend 700 MHz spectrometer equipped with a Prodigy TCI probe, controlled by Bruker Topspin software. For oxidation monitoring, 0.2 M cleaved peptide (PatE2K-2het-tryp) was reacted with 10 µM ArtGox in 10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM FMN and 5% D2O at °C. ^{1}H, ^{15}N HSQC spectra were recorded before the addition of enzyme, 5 min and 20 min afterwards, and then every 30 min. Six transients were recorded at 1536×80 points with spectral resolutions of 12.8 Hz and 26.6 Hz in the direct (^{1}H) and indirect (^{15}N) dimensions, respectively. ^{1}H, ^{15}N HSQC spectra were recorded at 10 °C for the unreacted substrate and the final product. Sequential assignment of the full-length peptide was transferred to the ^{1}H, ^{15}N HSQC spectrum at 10 °C and then to the corresponding spectrum at 20 °C. Raw spectral data were processed by Topspin and analysed with CCPN Analysis 2 [159].

4.3.4 Oxidation Reaction with Heterocyclized ITACITFCAYDGEK

100 µM PatE3KK (see section 3.3.2) was reacted with MicD (5 µM) in 100 mM Tris pH 8.0 supplemented with 50 mM NaCl, 10 mM $MgCl_2$, 10 mM ATP and 5 mM DTT for 16 h at 20 °C to introduce heterocycles. Following heterocyclization trypsin was added at 1:100 mass ratio and incubated for 2 h at 37 °C, and inactivated by heating the sample at 100 °C for 10 min. ArtGox (various concentrations) and FMN (200 µM) was then added and the reaction was carried out at 20 °C for 16 h. Reacted samples were subjected to ESI-MS on a AbSciex 5600 system, as described in section 3.3.6.

4.4 Results

4.4.1 Oxidation by ArtGox proceeds in an N-to-C order

To investigate the reaction order of ArtGox, its reaction with ^{15}N-labelled PatE2K-2het was monitored by ^{1}H, ^{15}N HSQC NMR (Figure 4.3). Having lost their amide proton, the heterocyclized residues (C47^{het}/Thn47 and C51^{het}/Thn51) do not themselves give cross peaks on these spectra. However, oxidation alters the chemical environment of neighbouring residues A46, I48, T49, F50, A52, Y53, D54 and G55, causing their resonances to shift. These changes can therefore be traced to indicate Thn47 and Thn51 oxidation. When the doubly hetereocyclized, full-length substrate (PatE2K-2het) was reacted with ArtG at 20 °C, changes took place in two distinguishable stages. The first period took around 12 h, demonstrated by the shifting of cross peaks corresponding to A46, I48, T49 and F50, residues close to Thn47. The second stage, which saw shifting of residues closer to C51 (F50, A52, Y53, D54 and G55), was complete in 60 h at 20 °C. These results showed that similar to the heterocyclase TruD, the oxidase ArtGox works in a distributive manner with an N-to-C order - preferentially catalysing the N-terminal thiazolines on different molecules before moving onto the C-terminal ones - in contrary to the C-to-N order shown by TruD.

4.4.2 Reaction with Leaderless Peptide Follows Same Order

Next, ^{15}N-labelled PatE2K-2het that has been digested by trypsin to remove the leader sequence was reacted with ArtGox, so as to determine whether the reaction order is dependent on leader-protein interaction. Similar to PatE2K-2het, the leaderless peptide was oxidised first on the N-terminal thiazoline (Figure 4.4), accompanied by chemical shifts of A46, I48, T49 and F50. Surprisingly the reaction rate was substantially faster than that between ArtGox and the full-length substrate, with the first stage being completed in 20 min. Rate of oxidation of the C51-derived thiazoline was also accelerated, and the reaction

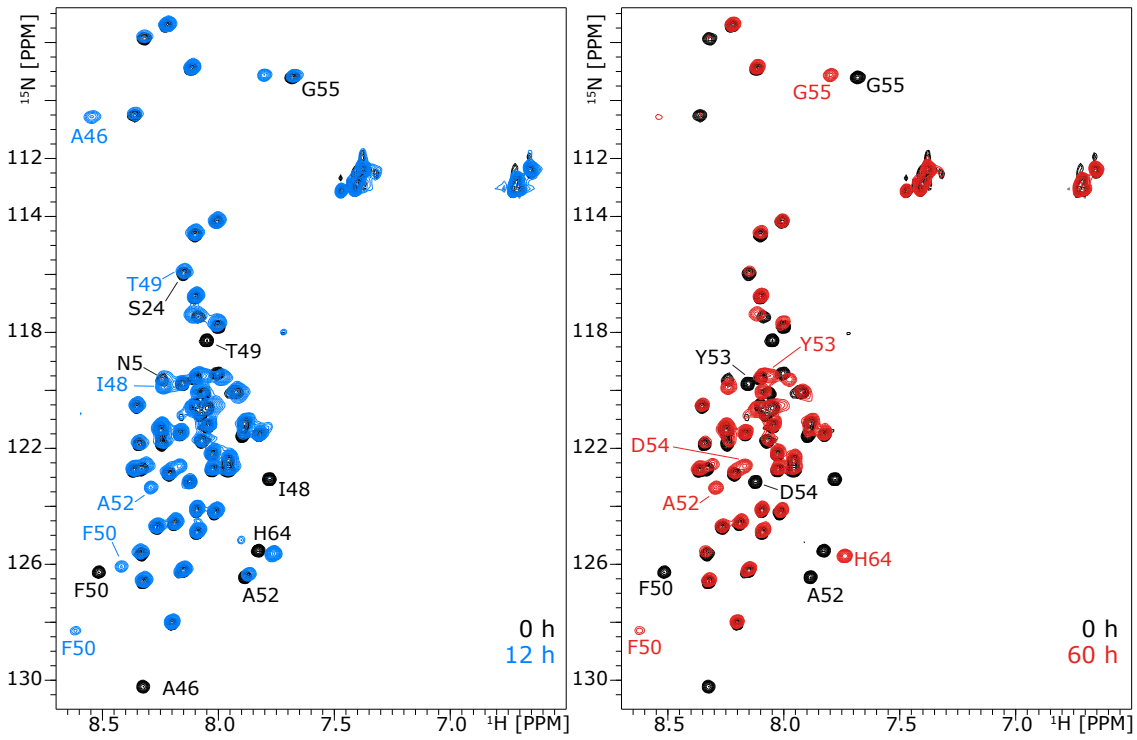

Figure 4.3: Oxidation Order of Doubly Heterocyclized PatE2K. ^{1}H, ^{15}N HSQC spectra of doubly heterocyclized PatE2K (200 µM) reacted with ArtGox (10 µM). The starting material is shown in black, while the 12 h intermediate and the 60 h final product are shown in blue and red, respectively. Cross-peaks are annotated with their residue sequences; the colour of a label corresponds to the spectrum where the cross-peak it designates first appears. Note that the cross-peaks for A46 after 12 h and 60 h are aliased in the 15N dimension; the actual 15N chemical shift for these signals is 133.8 ppm.

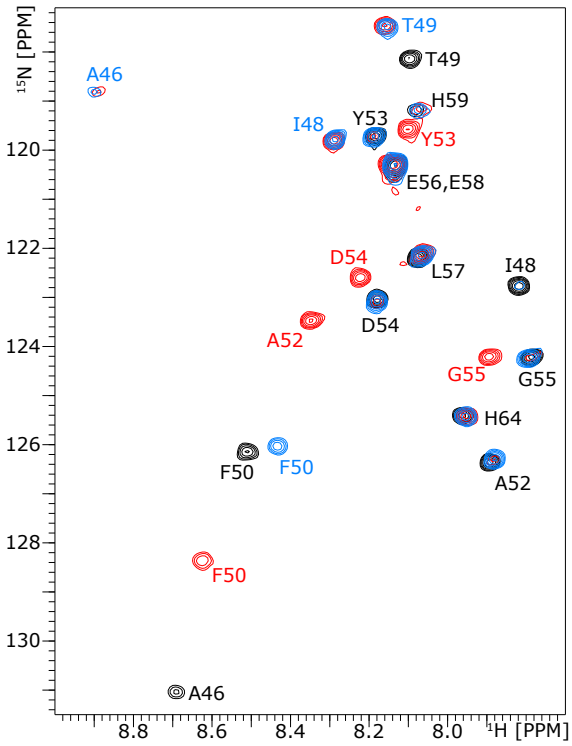

Figure 4.4: Oxidation Order of Doubly Heterocyclized Leaderless PatE2K. ^{1}H, ^{15}N HSQC spectra were recorded for doubly heterocyclised, leaderless PatE2K (black), and after it was reacted with ArtGox for 20 min (blue) and 16 h (red). Cross-peaks are annotated with their residue sequences as of the full-length substrate; the colour of a label corresponds to the spectrum where the cross-peak it designates first appears. Note the cross-peaks for A46 and G55 after 20 min and 16 h are aliased in the 15N dimension; the actual 15N chemical shift for these signals is 133.8 ppm and 109.3 ppm, respectively.

was finished within 16 h. The consistent N-to-C reaction order indicates the reaction order is intrinsic to ArtGox and the core substrate sequence (ITAC^{het}ITFC^{het}). Sequential assignments were transferred from the full-length peptide spectra to the leaderless peptide spectra (both taken at 20 °C), then to the leaderless peptide at 10 °C (Figure 4.5).

4.4.3 Oxidation of Oxazolines is Kinetically Hindered

As arthrospiramides discovered to date (Figure 4.2) contained only thiazoles, ArtG's ability to oxidise methyloxazolines was tested. 100 µM PatE3KK was reacted with MicD to introduce four heterocycles and subjected to proteolysis by trypsin to remove the leader and his-tag. The resulting peptide IT^{het}AC^{het}IT^{het}FC^{het}AYDGEK was then reacted was ArtGox and analysed by ESI-MS (Figure 4.6). 1 µM of ArtGox was able to oxidise the peptide twice - corresponding to the formation of two thiazoles. However, increasing the enzyme concentration up to 10 µM did not result in further oxidation within the time scale of the reaction (16 h). A third loss of 2 Da (1 Da in the doubly charged ion) was

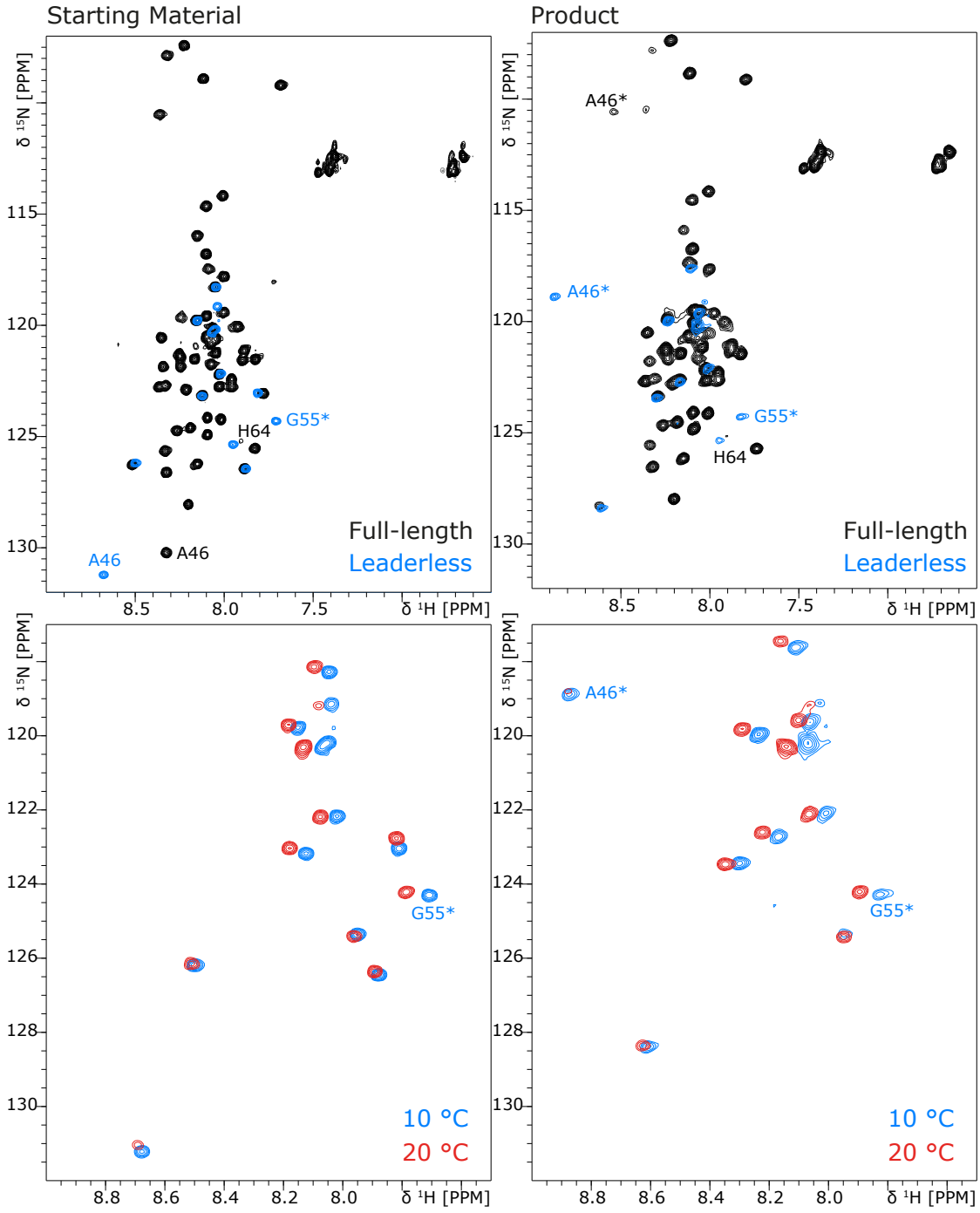

Figure 4.5: Sequential assignment of doubly heterocyclised, leaderless PatE2K and its oxidised product. Doubly heterocyclised full-length PatE and its oxidised product were assigned using triple resonance data. The sequential assignments were transferred to equivalent leaderless peptides at 10 °C. The assignments were then transferred to equivalent resonances at 20 °C. Cross-peaks marked with an asterisk are aliased in the 15N dimesion; the actual 15N chemical shifts for these signals are 133.8 ppm (A46) and 109.3 ppm (G55).

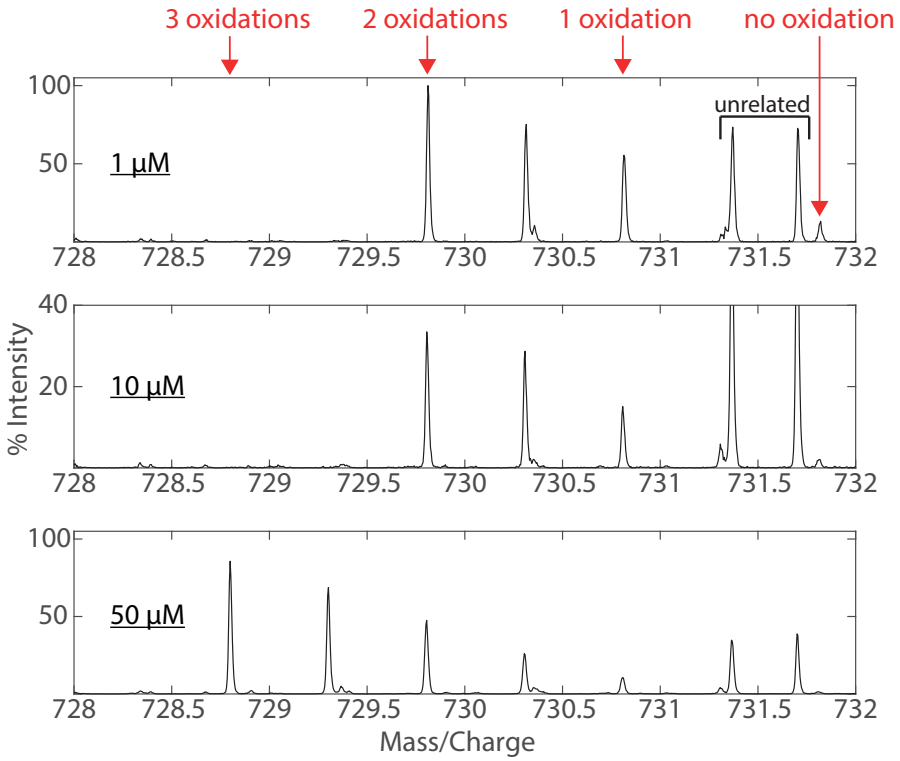

Figure 4.6: ESI-MS of IT^{het}AC^{het}IT^{het}FC^{het}AYDGEK reacted with ArtGox. PatE3KK (100 µM) was reacted with MicD, trypsinated and then reacted with ArtGox. Unoxidised mass of ITACITFCAYDGEK-4het ([M+2H]$^{2+}$)is expected to be 731.8 Da, while each oxidation results in a 2 Da mass loss (1 Da in the [M+2H]$^{2+}$ species). Note that the third isotope of the once oxidised peptide coincides with the unoxidised peptide, and so forth, and the less oxidised species are therefore likely overrepresented. The peaks labelled 'unrelated' were not generated from the peptide of interest, as they exhibit a different isotope pattern.

observed when ArtGox concentration was raised to 50 µM. Even at such high enzyme concentration, a fourth oxidation did not take place. These results suggest it is unlikely that this enzyme would catalyse methyloxazole formation on leaderless linear peptides under natural circumstances.

4.5 Discussion

The work described in this chapter showed that the cyanobactin oxidase ArtGox catalyses multiple thiazolines in a distributive manner, following an N-terminal to C-terminal order (C^{het}47 then C^{het}51), in an interesting contrast to that of heterocyclization seen in TruD and MicD (C51 then C47). This reversion of reaction order could provide a check point which ensures heterocyclization precedes oxidation. It is however also possible that the heterocyclase and oxidase work in tandem, and C51 is oxidised immediately after it is heterocyclized. Further experiments need to be carried out to determine whether C47

thiazoline is required for oxidation to initiate, or that C51 thiazoline alone is sufficient. In either case, the reaction order is attributed to intrinsic properties of ArtGox and the core peptide used, rather than the leader-mediated positioning, as removal of the leader peptide from the substrate did not affect the order of reaction.

It was also observed that leaderless peptides are processed at a higher rate compared to their full-length counterparts. A consistent discrepancy (50-fold increase in k_{cat}/K_m when leaderless peptide is used) was seen in the enzyme ThcOx when subjected to steady-state kinetics measurements [168]. Based on isothermal calorimetric data, Gao *et al.* [168] also reported that there was no interaction between ThcOx and the leader peptide, leaving it unexplained as to how the reaction is slowed down. The additional bulkiness of the leader sequence (38 residues) in PatE, as opposed to the leaderless peptides, could be one of the contributing factors. The large difference in reaction rates may suggest that proteolytic removal of the leader peptide by PatA occurs before oxidation in the native system.

It is unclear why the peptide clamps are conserved among cyanobactin oxidases (Figure 4.1), given they are not involved in leader binding. A similar observation was made of the peptide-modifying radical SAM enzyme SuiB [169], which contains a peptide clamp, but uses a different site for leader recognition. For ThcOx specifically, the N-terminal domain appears to play a stabilising role, as truncation of this domain resulted in the loss of the co-factor FMN and inactivation of the enzyme [168]. As natural product biosynthetic enzymes tend to be clustered on a genetic level, it is possible that lateral transfer of domains have arisen during evolution, leading to the fusion of RREs to enzymes that do not require them for substrate recognition. These examples demonstrate the importance of biochemical and biophysical characterisations of proteins of interest, and caution against assigning functions based on bioinformatic predications alone.

Chapter 5

Phosphate Chemistry of Cyanobactin Heterocyclization

5.1 Summary

YcaO-class heterocyclases have been proposed to operate via a kinase mechanism, where the peptide forms a hemiorthoamide intermediate which is phosphorylated, and the phosphate group is subsequently eliminated, yielding one each of heterocycle, ADP and inorganic phosphate. However, AMP and PP_i have also been observed in multiple studies. Specifically, an ^{18}O-labelling experiment showed that PP_i contained peptide-derived oxygen atoms. ATP analogues were employed to probe the mechanism, and it was shown that a cleavable bond between β,γ-phosphates is essential for reaction. Next, ATP/AMP-CPP metabolites and heterocycles were quantitated using HPLC-MS and radioactive methods, which demonstrated that ADP/AMP-CP production was correlated with heterocycle production, while AMP was generated from ADP hydrolysis. The phosphorylation theory was proposed that unify these observations, where the γ-phosphate dissociates from the β-phosphate in order to phosphorylate the hemiorthoamide, but then returns to the latter and form PP_i with it, simultaneously releasing AMP. In accordance with this theory PP_i was observed in place of P_i when AMP-CPP and ADP were added to the same reaction.

5.2 Introduction

Phosphoryl transfer forms the basis of fundamental aspects of biology such as DNA synthesis and ATP hydrolysis [170, 171]. As discussed in section 1.3.2, YcaO-catalysed heterocyclization is believed to proceed via a tetrahedral hemiorthoamide intermediate

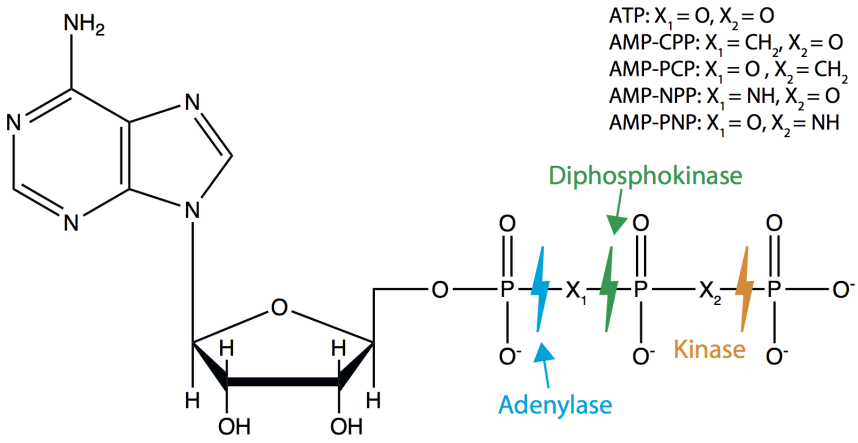

Figure 5.1: A schematic illustration of how adenylases, diphosphokinases and kinases cleave ATP.

formed when the side chain alcohol/thiol group of a Ser/Thr/Cys attacks the carbonyl carbon of its preceding amide bond. Following that, an ATP-derived phosphate group is attached to the former carbonyl oxygen to assist its elimination, but the exact identity of this electrophile is unknown. AMP, PP_i, ADP and P_i have all been observed as products of heterocyclization, giving possibilities that the reaction proceeds via a mechanism which will be referred to as adenylase, diphosphokinase or kinase here (Figure 5.1). In the adenlyase mechanism, AMP is used to activate the hemiorthoamide, whereas PP_i and P_i, respectively, are attached to the tetrahedral oxygen via the diphosphokinase and kinase routes. In this chapter the phosphate chemistry of the cyanobactin heterocyclase is studied in depth, to address these seemingly irreconcilable observations.

5.3 Materials and Methods

5.3.1 Protein Expression and Purification

LynD, LynD fusion, MicD, MicD fusions and OscD were purified as described in section 2.3.5.

5.3.2 Heterocyclization Reactions

ATP, ADP, AMP and various ATP analogues were reacted with heterocyclases in 100 mM Tris pH 8.0 supplemented with 150 mM NaCl, 20 mM $MgCl_2$ and 5 mM DTT at room temperature. Unless specified, the following concentrations of reactants were applied: 5 µM heterocyclase, 100 µM PatE, 5 mM nucleotide.

5.3.3 Phosphate versus Pyrophosphate Assay

EnzChek (Invitrogen) phosphate assay (see section 2.3.11) was carried out in the presence and absence of pyrophosphatase. Experiments were performed at 25 °C in 100 mM Tris pH 8.0 supplemented with 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT, 200 µM MESG and 1unit PNP (as supplied by the manufacturer).

For determination the products of MicD- and LynD-catalysed reactions, MicD (5 µM) was reacted with 20 µM PatE2K and 10 mM ATP or 10 mM ATP only, and LynD (5 µM) was reacted with 100 µPatE2K and 1 mM ATP or 1 mM ATP only. 10 µM *Mt*PPase was included in the reaction where indicated.

For the assay probing transphosphorylation, LynD (5 µM) was reacted with 100 µM PatE2K and 10 mM each of ATP, ADP, AMP-CPP or AMP-CPP + ADP. 10 µM *Mt*PPase was included in the reaction where indicated.

5.3.4 HPLC Assay

Heterocyclization of ITACITFCAYD by MicD fusion in the presence of ATP, ADP or AMP-CPP were carried out in HPLC assay buffer (100 mM Tris pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM TCEP). Reactant concentrations, reaction time and temperature varied and are specified individually. Reactions were terminated by the addition of two volumes of urea quench buffer (8 M urea, 20 mM Tris pH 8.0, 500 mM NaCl, 10 µM L-tryptophan), where L-tryptophan is used as an internal standard, and then incubated with 120 µM Ni-NTA agarose resin (ABT) placed in a Corning centrifuge tube filter (Sigma) for 30 min at room temperature. After centrifugation at 4000 g for 10 min the flow through was collected and applied to a EC 250/4.6 NUCLEODUR 300-5 C18 column (Thames Restek) connected to a G6130B Single Quad LC-MS instrument (Agilent Technologies). The column was pre-equilibrated in solvent A (5 mM Ammonium Bicarbonate pH 7.0), and after sample application a step-wise isocratic programme was run at a flow rate of 1 ml/min for a total of 11 min to separate the reactants and products: 80% solvent A, 20% solvent B (95% Acetonitrile) for 3 min, followed by 50% A, 50% B for 3 min, and finally 100% B for 5 min. Separation of linear and heterocyclized peptides was not achieved by this method and therefore different selected ion monitoring (SIM) channels were set up to give separate ion chromatograms of these peptides. The following three SIM channels (negative mode) were set up: (1) AMP or AMP-CP and ITACITFCAYD, (2) ATP or AMP-CPP and ITACITFCAYD-1het, (3) ADP and L-tryptophan. Ion intensity peaks were integrated using the Agilent ChemStation software, and corrected against the area of the L-tryptophan peak. Standard curves for the quantitation of AMP, ADP, ATP, AMP-CP, AMP-CPP and ITACITFCAYD was obtained by diluting each compound to various concentrations in HPLC assay buffer and adding two volumes of urea quench buffer, before

applying the samples to the HPLC-MS system.

Duplicated time course experiments were carried out with MicD Q21 9GA (20 µM) with ITACITFCAYD (50 µM) and ATP or AMP-CPP (10 µM). The reaction containing ATP was performed at 25 °C , with samples withdrawn from the reaction mix at 15 s, 30 s, 60 s, 120 s, 300s and 600 s after the start of the reaction, and quenched by being mixed with two volumes of urea quench buffer. The AMP-CPP reaction was carried out in separate aliquots and incubated at 30 °C; reactions were quenched after 15 min, 30 min, 60 min, 120 min and 240 min after initiation. Areas of ion intensity peaks (peptides, nucleotides and typtophan) were integrated and the plotted as a ratio of peptide or nucleotide over tryptophan, against time. Standard curves for ATP, ADP, AMP and AMP-PNP were employed to calculate the concentrations of these nucleotides. Due to difficulties in obtaining pure, singly heterocyclized peptide, the concentration of this compound was also obtained from normalised data. The AMP-CPP reaction was assumed to have reached completion by the end of the incubation period, and the readings at time 0 and 240 min, respectively, were used as minimum (0%) and maximum (100% or 10 µM), and a linear relationship between area ratio and peptide concentration was supposed to exist within this range. The same slope (area ratio/concentration) was then used to normalise data from the ATP reaction. AMP-CP concentrations were estimated based on normalised data, with the peak ratio value at time 0 set as 0 and the peak ratio value at 240 min as 100% (10 µM), assuming linearity within the range.

For determining the degree of heterocyclization, reactions were set up between 25 µM, 50 µM and 100 µM ITACITFCAYD and 1 mM AMP-CPP, catalysed by MicD Q21 9GA (5 µM). The reactions were incubated at room temperature for 16 h, and quenched with 2x volumes of urea quench buffer. The enzyme was removed, and the flow-through was subjected to HPLC-MS as described above, except that a different solvent gradient was run: 20%-60% B from 0 -10 min, 95% B from 10 -15 min. Masses corresponding to the unmodified, singly and doubly heterocyclized peptides were entered into SIM channels MSD2, 3 and 4, respectively.

5.4 Results

5.4.1 The Conundrum of ATP Consumption

A major differentiating factor between a kinase mechanism and adenylase/diphosphokinase mechanism lies in the phosphoester bond that is cleaved. Kinases transfer the γ-phosphate of ATP to another molecule, thereby requiring the cleavage of the bond between β- and γ-phosphates; the separation of AMP and PP_i, on the other hand, severs the bond between the α- and the β-phosphate (Figure 5.1). Four ATP analogues, two with a non-cleavable

α,β-bond, AMP-CPP and AMP-NPP, and two with a substituted β,γ-oxygen, AMP-PCP and AMP-PNP, were employed to probe the phosphate chemistry of cyanobactin heterocyclization. MicD is capable of heterocyclizing PatE2K (core sequence ITACITFC) four times when supplied with ATP (Figure 5.2A). When AMP-PCP/AMP-PNP was used in place of ATP, activity was not observed (Figure 5.2D, E). In contrast, heterocyclization was carried out when AMP-CPP and AMP-NPP were supplied as the phosphate donor, albeit at reduced rates (Figure 5.2B, C). AMP-CPP appeared to be a better substrate than AMP-NPP, although both allowed the reaction to take place. Since the peptide reacted with AMP-CPP/AMP-NPP but barely went beyond one heterocyclization, we considered whether the first reaction may have a different mechanism from the rest (one that produces pyrophosphate). To test this, PatE2K was reacted first with AMP-CPP for 16 h before the addition of AMP-PCP and 24 h further hours of incubation, in order to see if progression of further heterocyclization could be promoted (Figure 5.2F). However, the additional AMP-PCP did not make a difference to the reaction. These results agree with a kinase mechanism, where cleavage of the β,γ-phosphoester bond, but not that of the α,β-bond, is required for heterocyclization. The first reaction was faster than the rest, but it is likely down to kinetic properties rather than differences in phosphate chemistry. To show that the selectivity over nucleotides is not an exclusive characteristic of MicD, but common to the heterocyclase class of enzyme, PatE2K was reacted with the enzymes LynD and OscD in the presence of AMP-CPP and AMP-PCP (Figure 5.3). Both enzymes were capable of utilising AMP-CPP, but not AMP-PCP, to catalyse heterocyclization, echoing the behaviour of MicD.

If MicD-type heterocyclases operate via a kinase mechanism, when ATP is supplied instead of its analogues, it is expected to be broken down into ADP and inorganic phosphate, while the bond between the α- and β-phosphates remain intact. However, when measuring reaction rates of heterocyclases using the EnzChek phosphate assay, it was observed that including a pyrophosphatase (PPase) in the assay increased the readings. As the assay detects phosphate, a rational explanation would be that pyrophosphate was produced by the heterocyclization reaction and only became detectable after being broken down into phosphate by the pyrophosphate. This was first noticed when PatE2K was reacted with MicD fusion (Figure 2.9B). To rule out that a change in chemistry was induced by enzyme engineering, native MicD-catalysed reaction was subjected to the same analyses. MicD native was incubated with 20 µM PatE2K and 1 mM ATP for 600 s, without or without pyrophosphatase. In the presence of PPase, the increase in A360 was about three times higher than when PPase was not included, suggesting that approxiamately equal amounts of P_i and PP_i were produced during this period (PP_i breaks down into 2 P_i and therefore yields double the reading). Both rates were significantly higher than when the enzyme was incubated with ATP alone (Figure 5.4A), showing that PP_i was associated with the reaction rather than background hydrolysis. A consistent, and more pronounced difference was seen when LynD was employed to catalyse the reaction

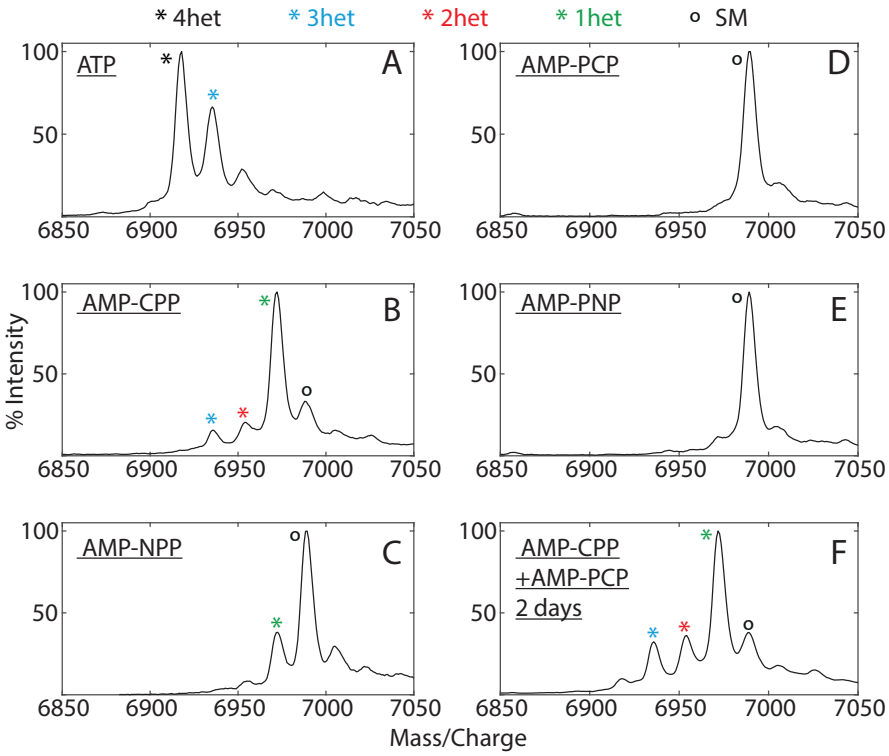

Figure 5.2: Usage of ATP Analogues by MicD. (A)-(E) 100 µM PatE2K (unreacted mass 6989 Da) was incubated with 5 µM MicD and 5 mM of varied nucleotides at 25 °C for 16 h. (F) 100 µM PatE2K was reacted with 5 µM MicD in the presence of 2 mM AMP-CPP for 16 h at 25 °C before 2 mM AMP-PCP was added and the sample was incubated for a further 40 h.

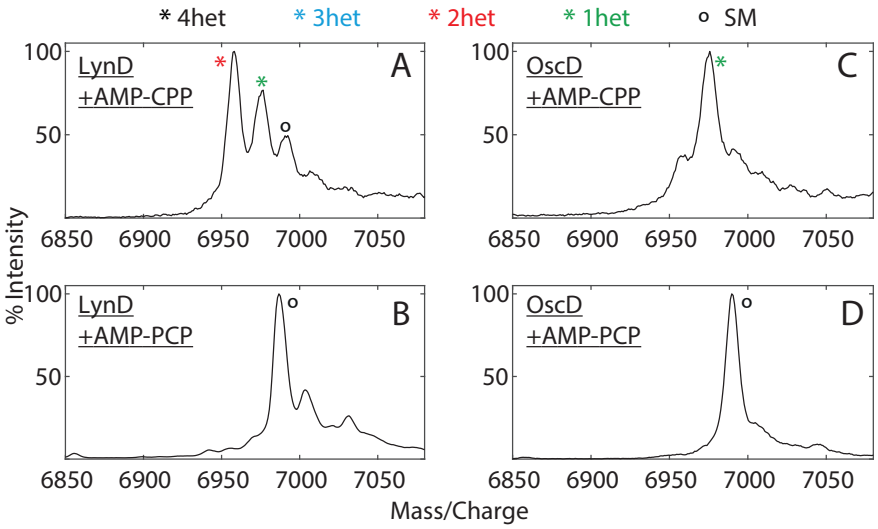

Figure 5.3: Usage of ATP Analogues by LynD and OscD. MALDI spectra of 100 µM PatE2K (unreacted mass 6989 Da) that has been incubated with 5 µM LynD (A and B) or OscD (C and D) and 5 mM of AMP-CPP or AMP-PCP at 25 °C for 16 h.

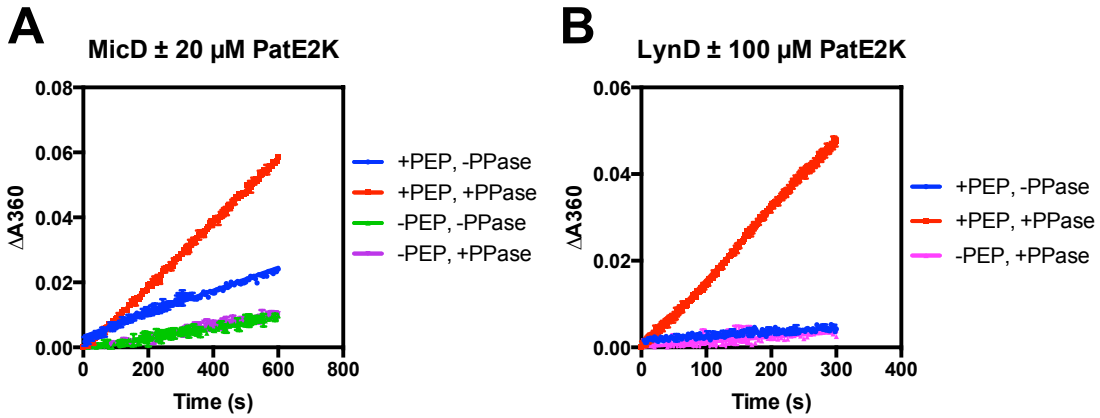

Figure 5.4: Pyrophosphate Production by Heterocyclases. Pyrophosphatase increases the amount of phosphate detectable following MicD (A) and LynD (B) reactions. PEP, peptide; PPase, pyrophosphatase.

between PatE2K and ATP (Figure 5.4B), with the addition of PPase boosting the reading for over 10 times. Without PPase, the LynD catalysed reaction yielded phosphate at a rate comparable to background (no peptide). It is therefore reasonable to conclude that PP_i is the main product of the heterocyclization reaction by LynD.

An alternative explanation for the PPase-related increase in A360 would be that a small amount of PP_i is produced by background ATP hydrolysis, but exerts an inhibitory effect on heterocyclization; breaking it down accelerates the reaction which produces more phosphate (in addition to the phosphate arisen from pyrophosphate hydrolysis). However, previous studies of LynD and TruD observed AMP and PP_i as reaction products [64, 65]. Specifically, the LynD study [64] showed, via ^{18}O-labelling of the carbonyl carbon preceding the cysteine to be heterocyclized, incorporation of ^{18}O into pyrophosphate, meaning that PP_i did not arise purely from background ATP hydrolysis.

The apparent contradiction between the detection of pyrophosphate and the need for a cleavable β,γ-phosphoester bond suggested that MicD-type heterocyclases may give rise to a novel, unusual mechanism.

5.4.2 Nuclease and Heterocyclase Activities of MicD

Having demonstrated that AMP-CPP/AMP-NPP, but not AMP-PCP/AMP-PNP could be used by MicD to produce heterocycles, experiments were carried out to show ATP is indeed cleaved to yield ADP and P_i. An LC-MS assay was developed to monitor the concentration of ATP and its metabolites ADP and AMP, as well as the peptide ITACITFCAYD and its heterocyclized forms (Figure 5.5). Nucleotide were separated by HPLC, and their masses were entered into different SIM channels to facilitate peak

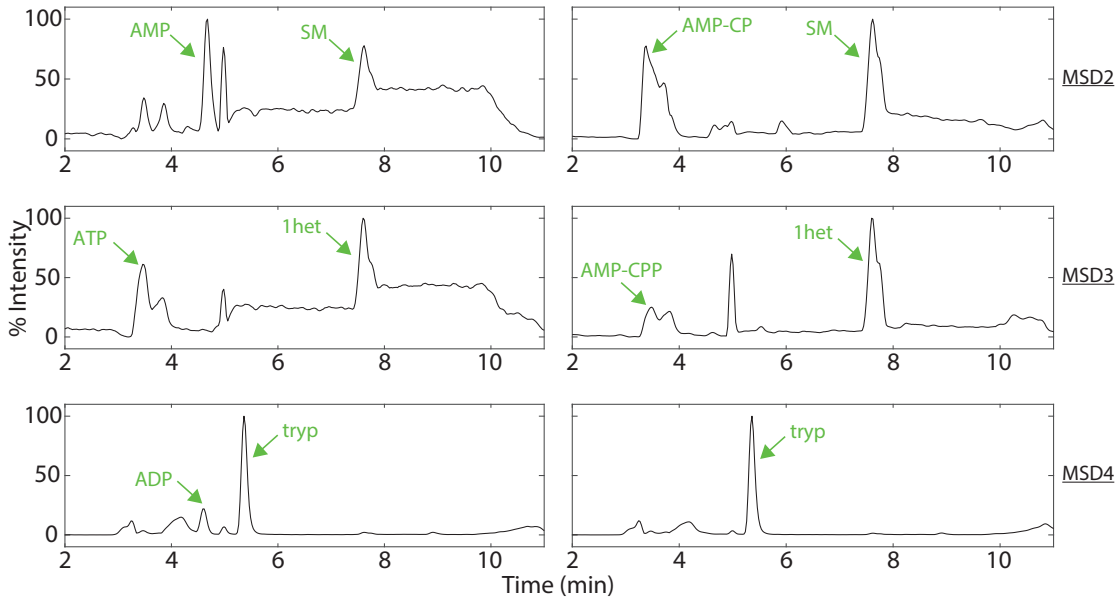

Figure 5.5: HPLC-MS Chromatograms of ATP/AMP-CPP Metabolites and ITACITFCAYD. Example chromatogram showing peaks for ATP, ADP, AMP, AMP-CPP, AMP-CP, the peptide ITACITFCAYD and its singly heterocyclized form, and L-tryptophan. 1het, singly heterocyclized peptide; tryp, L-tryptophan.

integration. The unreacted peptide was not separable using the current method from its singly heterocyclized product, and their quantitation relied on using different selective mass detectors (MSD). Separation of peptides was possible using acidic solvents (such as formic acid) but the nucleotides did not ionise to sufficient intensities under those conditions. In addition, azoline rings risk re-opening at low pH values. Tryptophan was included in the reaction quenching buffer to provide an internal standard for the assay.

As MicD is distributive with a strong preference for C8 of the peptide, all peptides are expected to be modified once at the same position (C8). It was experimentally validated that even when the nucleotide concentration far exceeds that of the peptide (multiple turnovers condition), singly heterocyclized peptide was exclusively produced as the starting material was consumed (Figure 5.6). Therefore the number of azolines were assumed to be equal to the amount of the singly heterocyclized peptide during the quantitation experiments. 20 µM MicD Q21 9GA was reacted with 10 µM ATP and 50 µM peptide, to create a pseudo-single turnover scenario; The reaction was quenched at 5, 10, 30 and 60 min by adding two volumes of urea quench buffer and subjected to LC-MS. The mass peaks were integrated and normalised against tryptophan, then plotted against time (Figure 5.7). The reaction appeared to have finished within the first 5 min, with concentration of the singly heterocyclized staying unchanged over the rest of the time course (Figure 5.7A, red circles). ATP (orange squares) was converted into AMP (green squares). In contrast, when ADP was added in place of ATP, only a residual level of 1het peptides were seen (Figure 5.7B), which could be attributed to noises in the mass

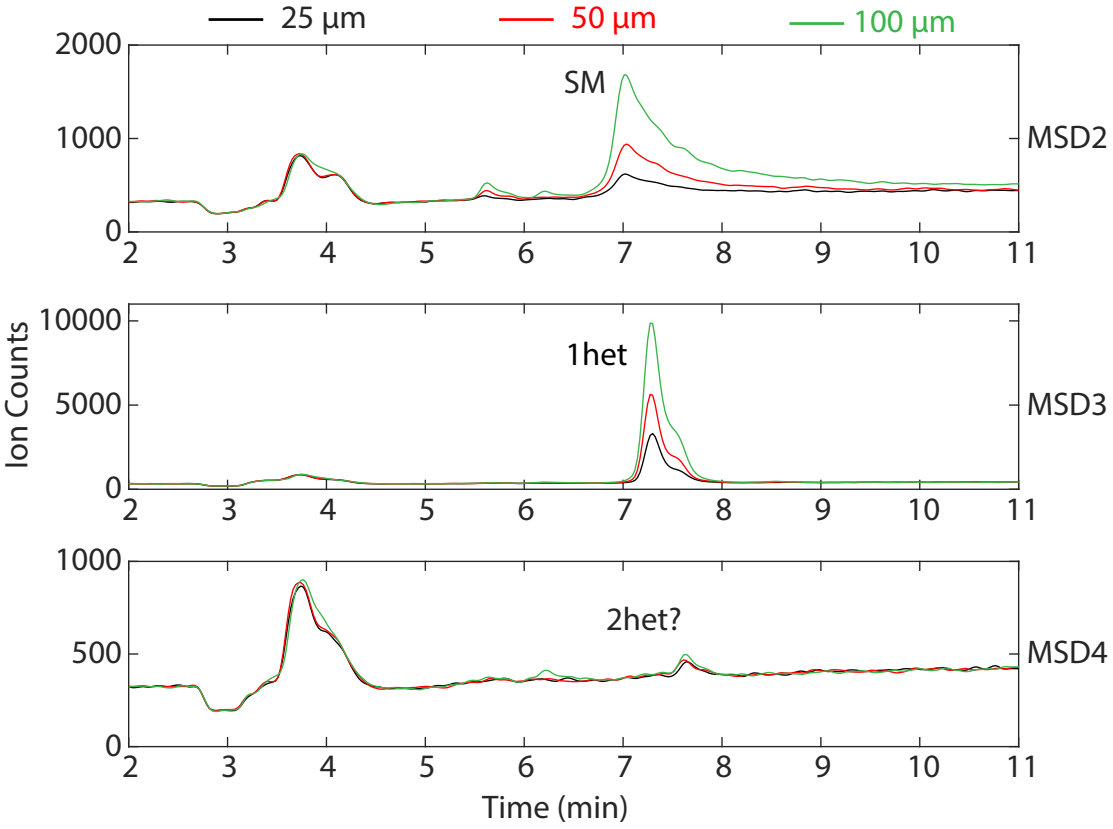

Figure 5.6: HPLC-MS Chromatograms showing masses corresponding to unmodified and heterocyclized ITACITFCAYD. 25 µM (black), 50 µM (red) and 100 µ(green) peptide was reacted with 1 mM AMP-CPP for 16 h prior to analysis. MSD2, unmodified mass; MSD3, singly heterocyclized (1het) mass; MSD3, doubly heterocyclized (2het) mass.

spectra. Interestingly, ADP was broken down to AMP gradually, although this process was not coupled with heterocyclization. Next, reactions were set up using AMP-CPP as phosphate donor, and the level of the degradation product AMP-CP was measured alongside the peptide. An increase in both was seen over the course of 60 min (Figure 5.7C), with the concentration of singly heterocyclized peptide approaching that of the ATP-reaction (Figure 5.7D) by the end of the incubation period.

These data suggested both ATP and AMP-CPP are used as substrates by the heterocyclase. ADP appeared to be hydrolysed into AMP, possibly in an enzyme-dependent way. In section 5.4.3, the reaction with ATP is examined at shorter time points after initiation, to reveal details about earlier stages of the reaction, and the reaction with AMP-CPP was monitored for a longer period of time. A method of quantitation of peptides and nucleotides was also developed to further facilitate the understanding of the chemistries involved in heterocyclization.

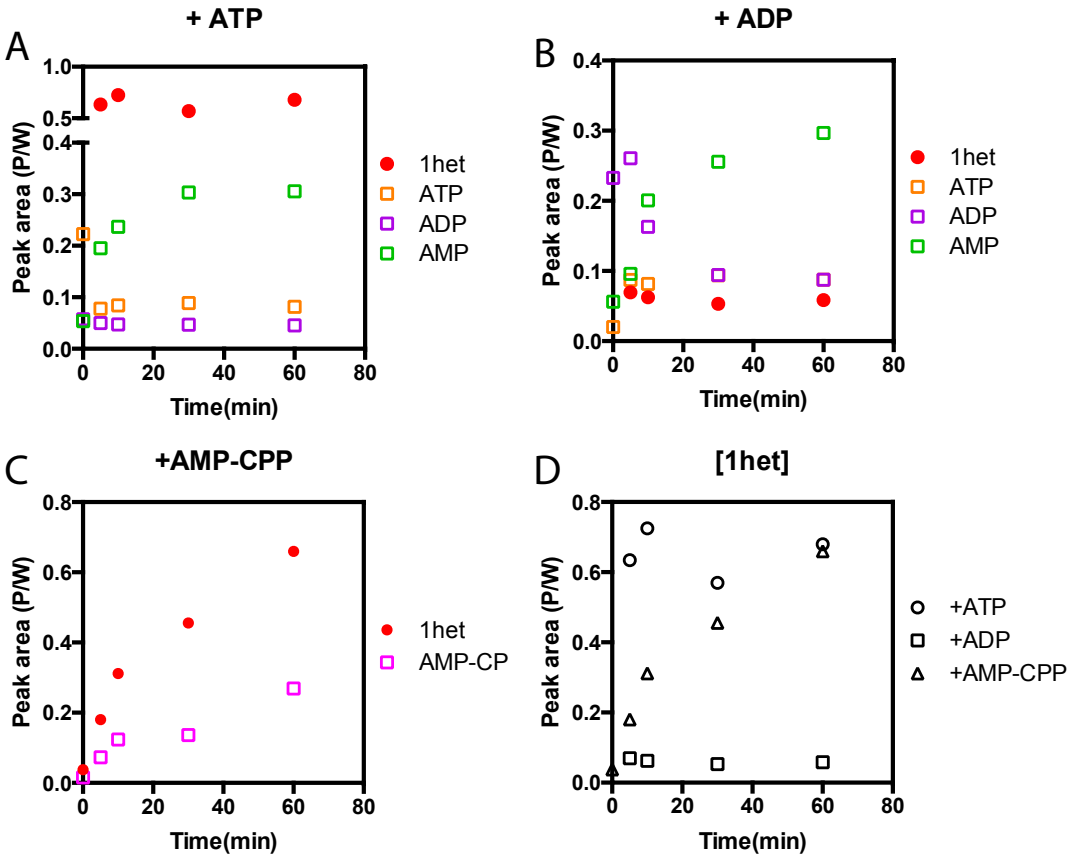

Figure 5.7: Association Between ATP, ADP and AMP-CPP Usage by MicD and Heterocyclization. ITACITFCAYD (50 µM) was reacted with 20 µM MicD Q21 9GA in the presence of 10 µM ATP, ADP or AMP-CPP for 5 min, 10 min, 30 min and 60 min. LC-MS peaks were integrated and corrected against an internal standard (tryptophan). P/W, area of the peak of interest over the area of tryptophan peak; 1het, singly heterocyclized ITACITFCAYD.

5.4.3 Quantitation of Nucleotides and Heterocycles

Based on the results (Figure 5.7) showing that ATP and AMP-CPP are broken down as heterocycles form, a method of quantitation of nucleotides and heterocycles were sought so as to determine the stoichiometry between products. Standard curves were established for ATP, ADP, AMP and AMP-CPP, (Figure 5.8A), as well as for the peptide ITACITFCAYD (Figure 5.8B). Reactions were carried out between 20 µM MicD Q21 9GA, 50 µM peptide and 10 µATP or AMP-PCP. The ATP reaction was conducted at 25 °C for a total of 10 min, whereas the AMP-PCP reaction was incubated at 30 °C for 4 h. Consistent with previous observations (Figure 5.7), when ATP is used as phosphate donor, heterocyclization occurred rapidly (finished within the first 200 s) and plateaued afterwards (Figure 5.8C). A rapid but temporary accumulation of ADP was observed (peaked at around 100 s), which subsequently diminished. In contrast, AMP concentration increased over the entire 600 s period, even after ATP had been depleted (peaks became too small to integrate after 300 s), and concurred with ADP hydrolysis. Due to sensitivity limits of the HPLC-MS instrument, the reactions were not carried out under strict single turnover conditions. Concentrations of AMP, ATP and the singly heterocyclized peptide were fitted to a first-order association/decay equation to demonstrate the trend.

AMP-CPP, on the other hand, was degraded over a much longer period of time (Figure 5.8D), even though the incubation temperature was higher (30 °C) than that of the ATP reaction (25 °C). As the metabolite AMP-CP could not be further converted to AMP, it accumulated over the course of the reaction. It was assumed that its concentration of it at 240 min was 10 µM (while that at time 0 was 0), based on the observation that AMP-CPP had been exhausted by that time, and that within this range the relationship between concentration and ion intensity was linear (Figure 5.8A). Concentrations of AMP-CP over the course of the reaction was approximated based on this simplified calibration curve. Similarly, quantitation of the singly heterocyclized peptide was based on the assumption that it has reached 10 µM at the end of the reaction (Figure 5.8B). This was not an accurate calibration, which was hindered by difficulties in achieving and purifying singly heterocyclized peptides, but allowed comparisons of reaction rates. Under the premise that 10 µM of heterocyclized peptides were produced, it appeared that the number of heterocyclized peptide was nearly identical to that of AMP-CP throughout the reaction (Figure 5.8D), indicating that the breakdown of AMP-CPP and the heterocyclization of peptide was tightly coupled, with a 1:1 stoichiometry. Had the concentration of the singly heterocyclized peptide been different (e.g. 8 µM heterocycles were made in total), there would still be a linear, albeit not equivalent, relationship between the amount of AMP-CP and heterocycles, in coherence with a phosphorylation-related mechanism. In such case, background hydrolysis of AMP-CPP could have contributed to the discrepancy. Regardless, the data clearly demonstrated that AMP-CPP was broken down into AMP-CP as heterocyclization occurs, proving that AMP-CPP is a genuine substrate, in agreement

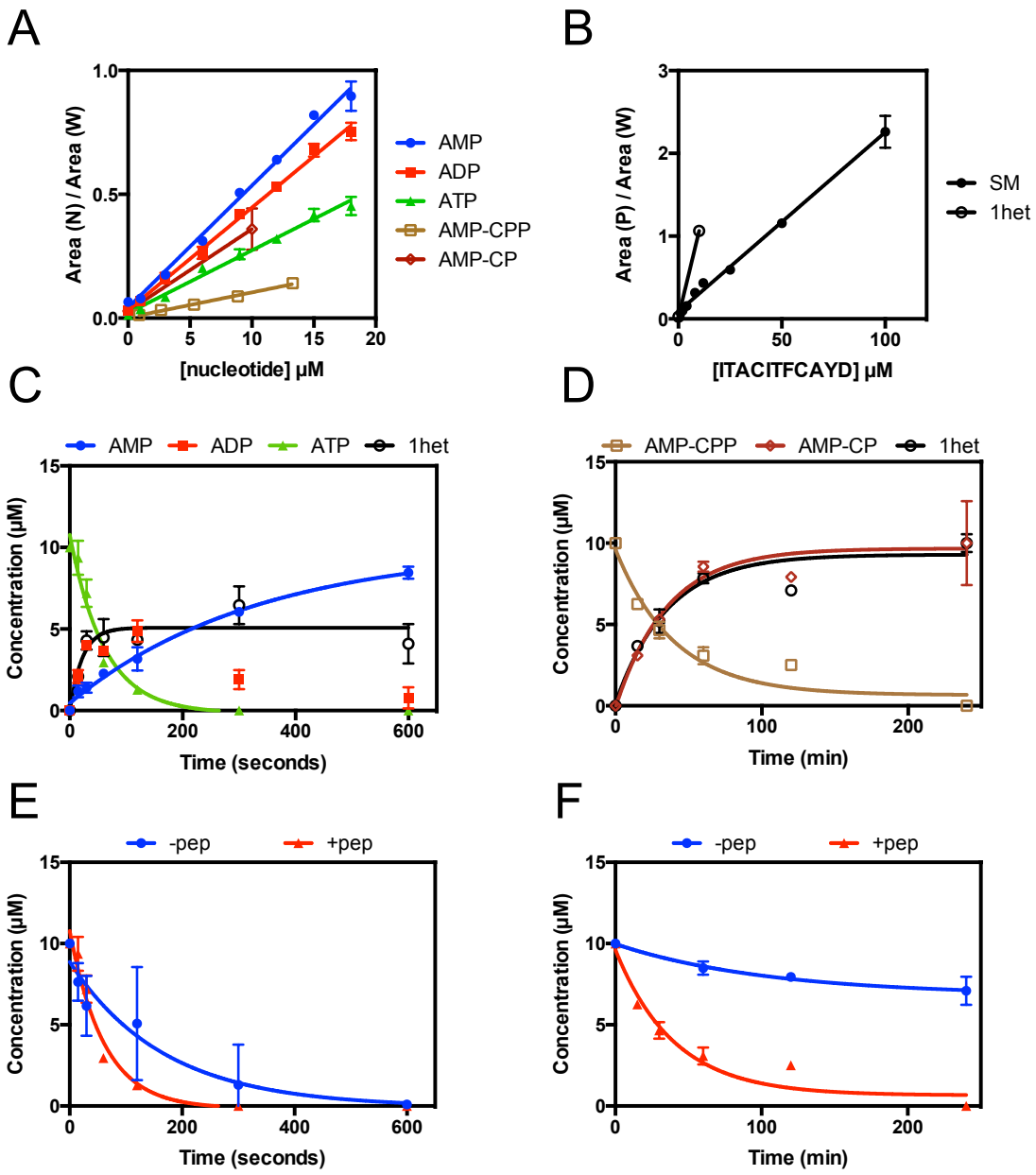

Figure 5.8: HPLC Quantitation of Nucleotides and Heterocycles. (A) Calibration curves for HPLC quantitation of the nucleotides ATP, ADP, AMP, AMP-CPP and AMP-CP. (B) Calibration curve for the peptide ITACITFCAYD and its singly heterocyclized derivative. Y axes represent the ratio between the area of the ion intensity peak and that of the tryptophan peak. (C) Progression of heterocyclization when the peptide was reacted with ATP. (D) Concentrations of metabolites when AMP-CPP was used as phosphate donor. (E) ATP degradation over time when MicD was incubated with ATP with and without the peptide. (F) AMP-CPP degradation when MicD was incubated with the nucleotide with and without the peptide. Data in (A) and (B) were fitted to linear regressions; Data in (C)-(F) were fitted to one-phase association/decay regressions.

with the observation of dehydrations by mass spectrometry (Figure 5.2) when this nucleotide is supplied as phosphate donor.

Even without knowing the exact number of heterocycles produced, it was clear that more (above twice the amount) was made during the AMP-CPP reaction (Figure 5.8D) compared with the ATP reaction (Figure 5.8C). ATP appears more prone to futile consumption, as when the peptide is absent ATP is hydrolysed not only faster than AMP-CPP in absolute terms, but the rate is more comparable to that observed during the reaction (Figure 5.8E, F).

The degradation pattern of ATP was iterated by a radioactive assay utilising α-^{32}P-labelled ATP (Figure 5.9), kindly carried out by Dr Clarissa Czekster (University of St Andrews). Native MicD (20 µM) was reacted with 0.25 µM ATP and a saturating amount of PatE2K (150 µM), and quenched with a final concentration of 62.5 mM EDTA at discontinuous time points (15 s to 1800 s). ATP was consumed at a rate highly similar to that seen in the HPLC experiment, with nearly 90% being degraded within the first 300 s (Figure 5.9B); ADP was produced, with a peak concentration of 0.134 µM at 120 s, and subsequently broken down into AMP. The increase in AMP (0.1 µM) approximately equalled the decrease of ADP (0.09 µM) between the 300 s and 1800 s time points, as was expected if AMP derived primarily from ADP hydrolysis. In the meanwhile, background ATP hydrolysis was measured (Figure 5.9B) by omitting PatE2K from the reaction. The overall ATP consumption differed little from the reaction when the peptide was present. However, a smaller amount of ADP accumulated (peaked at about 0.075 µM at 60 s), and AMP levels rose much more rapidly (nearly 90% AMP at 600 s). One possible explanation for the difference is that ADP is made from ATP at a slower rate compared to when the peptide is present, as the hemiorthoamide is not available to receive to the γ-phosphate from ATP now. Alternatively, ADP hydrolysis could be faster in the reaction shown in Figure 5.9A, although it is difficult to see how the presence of peptide would make a difference in that process. Lastly, the surplus of AMP could have been made directly from ATP, alongside PP_i - this would however contradict the observation that more PP_i is made when the peptide is present than when it is absent (Figure 5.4). On the current time scale (earliest point at 15 s after reaction started), a definite lag in AMP production behind that of ADP was not identified. It is therefore uncertain whether AMP was made exclusively from ADP hydrolysis, or originated from both ATP and ADP.

Taken together, the results descried in this section confirmed that MicD is capable of heterocyclizing peptide via a process that involves the transfer of a γ-phosphate group, and provided a convincing account for the generation of AMP that complies with this mechanism. The discovery of ADP hydrolysis into AMP, however, further compounded the author with regard to the origin of PP_i, which was originally thought to have derived from the degradation of ATP into AMP and PP_i. In addition, the observation that ATP was a better substrate than AMP-CPP in terms of reaction rate (Figure 5.8C, D), but worse in

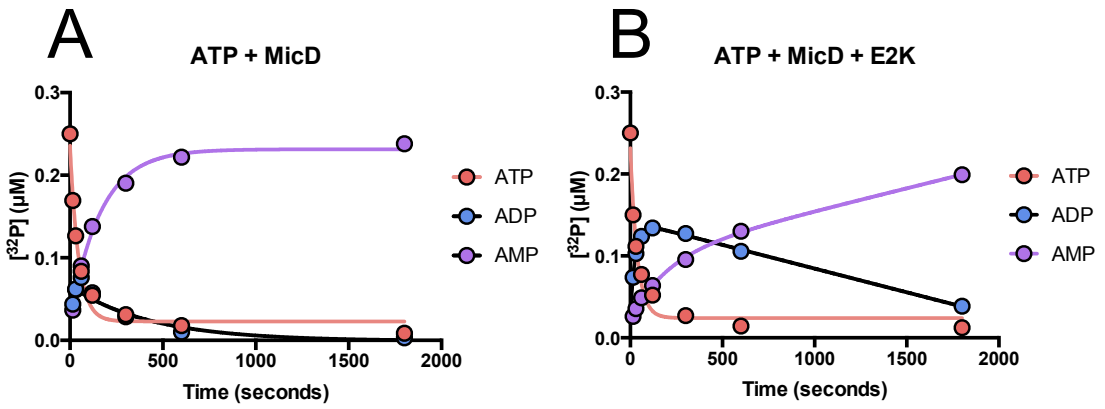

Figure 5.9: ^{32}P-labelled ATP broken down into ADP and AMP by MicD. (A) Degradation of ATP by MicD alone. (B) Breakdown pattern of ATP when PatE2K was present in the reaction mixture. The experiment was carried out by Dr Clarissa Czekster.

economy (fewer heterocycles produced using 10 µM ATP than 10 µM AMP-CPP) was an interesting one. The rate difference was intuitive, as substitution of the α,β-oxygen with the less electronegative carbon atom reduces the reactivity of the molecule. The change in bond angle (from 120° to 109.5° - may vary due to magnesium co-ordination) could also affect how well the nucleotide binds to MicD. The apparent waste of ATP might be a mere side effect of its high reactivity, but it is not impossible that background hydrolysis of ATP contributes to the reaction in a constructive way.

5.4.4 The Effect of ATP Metabolites on Heterocyclization

To test the effect of ATP metabolites (ADP, AMP, P_i or PP_i) on heterocyclization, AMP-CPP (2 mM), which is not naturally broken down into ADP , AMP or PP_i, was used as substrate. 2 mM ADP, AMP, phosphate (NaH_2PO_4) and pyrophosphate ($Na_4P_2O_7$) were supplemented individually to the reaction. After 16 h of reaction at 20 °C, the reaction with ATP as a substrate yielded a substantial amount of triply and quadruply heterocyclized peptides (Figure 5.10A), whereas when AMP-CPP was used the reaction went at a reduced rate, with the majority product being singly heterocyclized (Figure 5.10D). None of the additive compounds were able to accelerate the AMP-CPP reaction. Conversely, P_i and PP_i both had an inhibitory effect. It is tempting to relate this back to the EnzChek assay (Figure 5.4) where the addition of pyrophosphatase increase the apparent rate of reaction, and conclude that the breakdown of PP_i expedites the progression of heterocyclization; however, as P_i was also shown to hinder the reaction, the alleviation of inhibition should be minimal.

In conclusion, ATP hydrolysis does not contribute positively to heterocyclization. Instead, accumulation of P_i and PP_i as the reaction progresses, either as a result of the

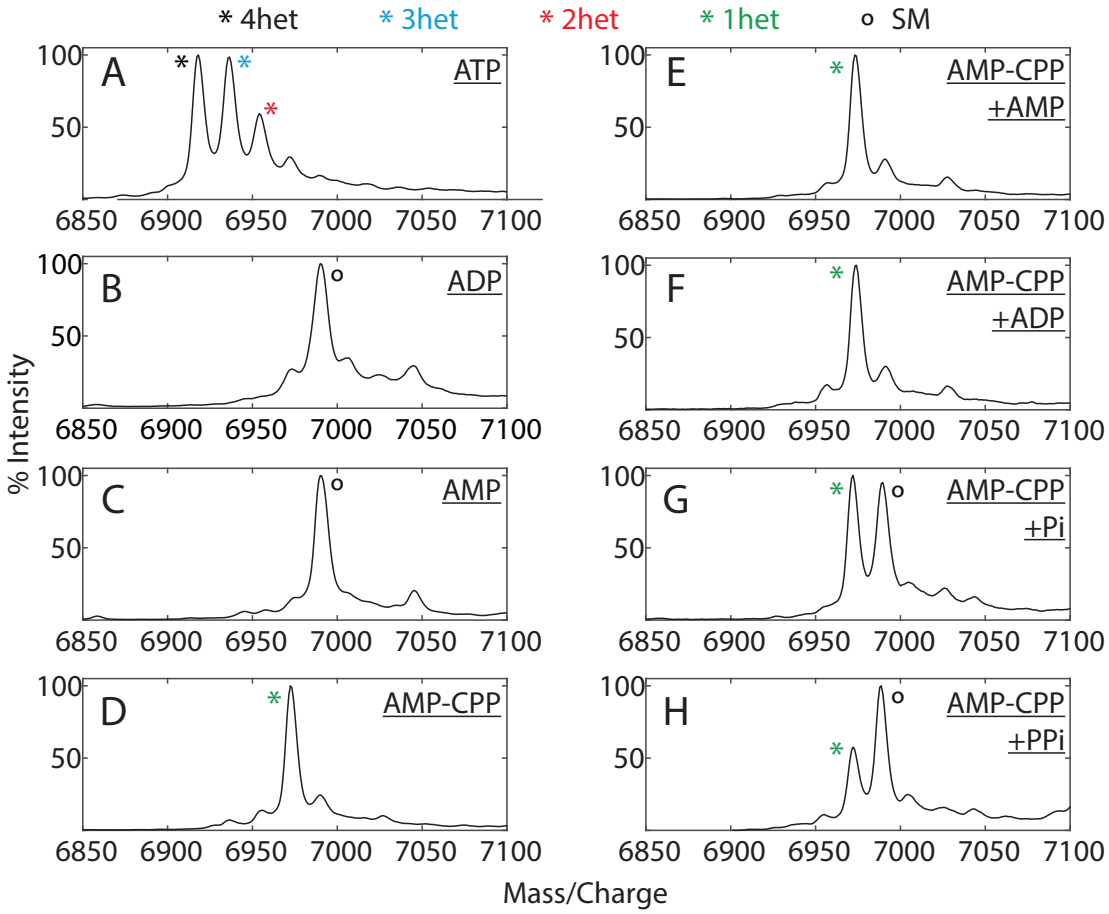

Figure 5.10: The Effect of ATP Metabolites on Heterocyclization. MicD (5 µM) was reacted with 100 µM PatE2K and (A) ATP, (B) ADP, (C) AMP, (D) AMP-CPP, (E) AMP-CPP + AMP, (F) AMP-CPP + ADP, (G) AMP-CPP + NaH_2PO_4 and (H) AMP-CPP + $Na_4P_2O_7$ in 100 mM Tris pH 8.0 supplemented with 150 mM NaCl, 10 mM $MgCl_2$ and 5 mM DTT. The concentration of each nucleotide/P_i/PP_i was 2 mM.

productive (heterocyclization) or the side reaction (hydrolysis), will likely cause product inhibition and slows down further catalysis.

5.4.5 A Unifying Mechanism

Having shown that AMP could be made from ADP hydrolysis (section 5.4.3), the origin of PP_i, which appears when the heterocyclase is incubated with both peptide and nucleotide, remains unresolved. An interesting observation was made when different nucleotides were incubated with PatE2K and LynD (Figure 5.11). The levels of P_i and PP_i produced during each reaction was monitored using the EnzChek assay. When 10 mM ATP was included in the reaction, a large discrepancy in ΔA360 was seen between the samples that contained pyrophosphatase and those that did not (Figure 5.11A), consistent with previous observations (Figure 5.4B). In contrast, when AMP-CPP was used instead of ATP, the

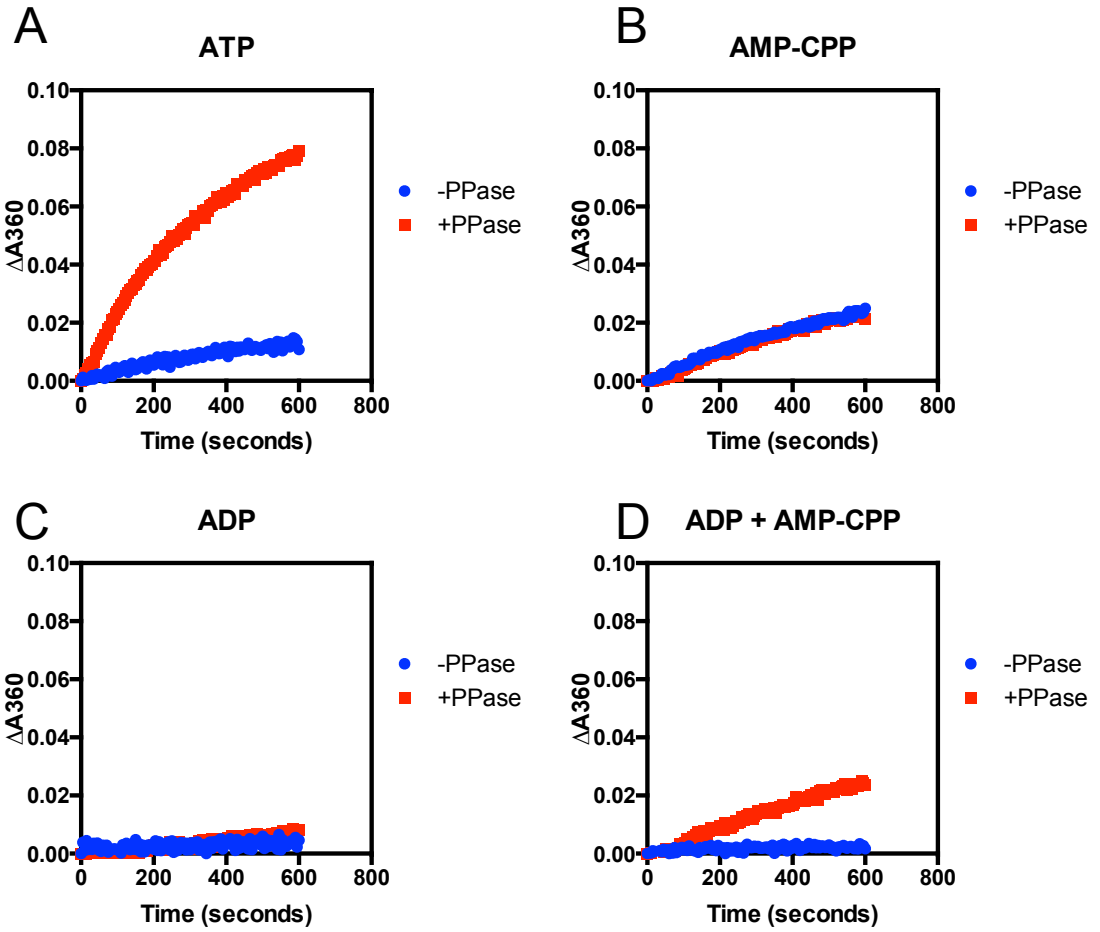

Figure 5.11: EnzChek Assay comparing the level of inorganic pyrophosphate during LynD (5 µM) reaction with PatE2K (100 µM) and 10 mM of (A) ATP, (B) AMP-CPP, (C) ADP and (D) ADP + AMP-CPP. Experiments were carried out in duplicates. ΔA360, changes in absorbance at 360 nm. PPase, inorganic pyrophosphatase.

reading remained the same (ΔA360=0.022 over 600 s) with or without the addition of pyrophosphatase (Figure 5.11B), indicating that only phosphate was produced during the reaction. Incubation with ADP gave rise to residual level readings only (Figure 5.11C), as it is not a substrate of heterocyclization. An interesting result was seen when AMP-CPP and ADP were both incubated with LynD and PatE2K (Figure 5.11D). The rate of phosphate release without added PPase was suppressed (compared to Figure 5.11B), suggesting that very little inorganic phosphate was produced. Upon the addition of PPase, phosphate release resumed to a level similar to when AMP-CPP alone was used as substrate (ΔA360=0.023 over 600 s). As neither AMP-CPP nor ADP could be hydrolysed to yield pyrophosphate, it must have arisen from a different mechanism. It should be noted that this level of increase in A360 needs 14.6 µM P_i to be reacted with MESG, but only 7.3 µM PP_i. Therefore, the consumption rate of AMP-CPP was around half of what it was when ADP was not supplemented.

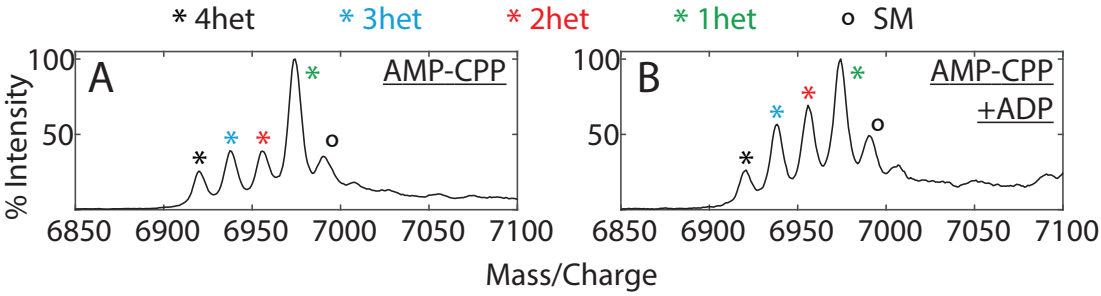

Figure 5.12: MALDI-MS analyses of PatE2K (100 µM) that have been reacted with MicD (5 µM) and (A)10 mM AMP-CPP or (B) 10 mM AMP-CPP + 10 mM ADP.

The rate of heterocyclization with AMP-CPP as substrate, with and without the addition of ADP, was then measured. Data from Figure 5.10 (D and F) showed that adding 2 mM ADP to 2 mM AMP-CPP did not change the level of heterocyclization of PatE by MicD after 16 h of reaction, compared to when AMP-CPP alone was added. After raising the concentrations to match the EnzChek assay (10 mM AMP-CPP plus 10 mM ADP), at which level PP_i appeared to be made (Figure 5.11), the reaction was subjected to MALDI-MS analysis. The overall pattern was still the same (Figure 5.12) as when 10 mM AMP-CPP only was supplied, with the singly heterocyclized peptide being the dominant species. The intensities of the peaks corresponding to multiple dehydrations increased relative to that of the singly heterocyclized species, but so did the peak for the unmodified peptide. These increases were therefore at least partly due to variations in ionisation strengths, rather than an actual change of the degree of heterocyclization. It might be interesting to carry out a quantitative study on a shorter time scale to show whether the change in phosphate chemistry indicated by Figure 5.11 alters reaction rate.

Another unexpected, but interesting observation was the frequent appearance of a peak at around 7106 Da, when PatE2K that have been reacted with AMP-CPP was subjected to MADLI-MS (Figure 5.13). This peak does not match that of PatE2K with a phosphate adduct and was originally believed to be enzyme-derived. However, it was present only when AMP-CPP (but not ATP) was supplied as phosphate donor, regardless of the enzyme used, and hence noteworthy to future investigators.

Based on our observations we devised a hypothesis (Figure 5.14) for pyrophosphate formation when AMP-CPP and ADP are incubated with PatE2K and LynD: first, the γ-phosphate of AMP-CPP is used to phosphorylate the precursor peptide; this phosphate, instead of being hydrolysed, is redirected to the β-phosphate of ADP, yielding PP_i and AMP. For AMP-CPP, this process could take longer than the hydrolysis of the phosphorylated hemiorthoamide, as ADP needs to replace AMP-CP, while for ATP the delay would be smaller, with ADP already optimally located after the first phosphorylation. Whether or not it is mechanistically beneficial for heterocyclization, this reaction (a form of transphosphorylation) explains not only the presence of PP_i, but more

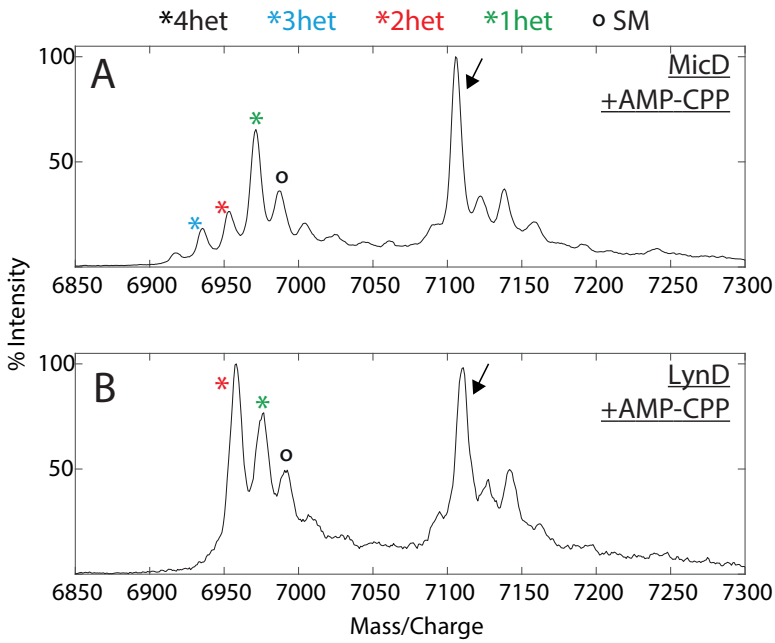

Figure 5.13: MALDI-MS data from (A) Figure 5.2B and (B) Figure 5.3A replotted to show a wider mass range. The black arrows point to the unknown peak that frequently comes up in AMP-CPP reactions.

importantly those that contain a peptide derived oxygen (shown via ^{18}O-labelling, [64]). It provides a unifying account that is consistent with all previous reports that showed ADP+P_i and/or AMP+PP_i as products of YcaO-catalysed azoline formation.

5.5 Discussion

The work described in this chapter dissected the ATP requirements of cyanobactin heterocyclization and analysed it in details. The experiments with ATP analogues (Figure 5.2) strongly suggest that the heterocyclase employs a kinase mechanism. This is further supported by the HPLC-MS results (Figure 5.8) and the radioactive assay results (Figure 5.9) which show the production of ADP, and its subsequently degradation into AMP.

With the origin of AMP explained, the only observed metabolites that could not have been generated by the kinase followed by ADP-hydrolysis mechanism was inorganic pyrophosphate. A transphosphorylation hypothesis was therefore proposed, which provides a plausible account for the presence of PP_i following heterocyclization. Current evidence suggests this could indeed happen: the combination of AMP-CPP and ADP, neither of which can be broken down to yield PP_i on its own, led to the observation of PP_i when incubated with LynD and PatE2K (Figure 5.11). It might be worth employing a more direct method, such as phosphate NMR (previously used in the study of TruD to

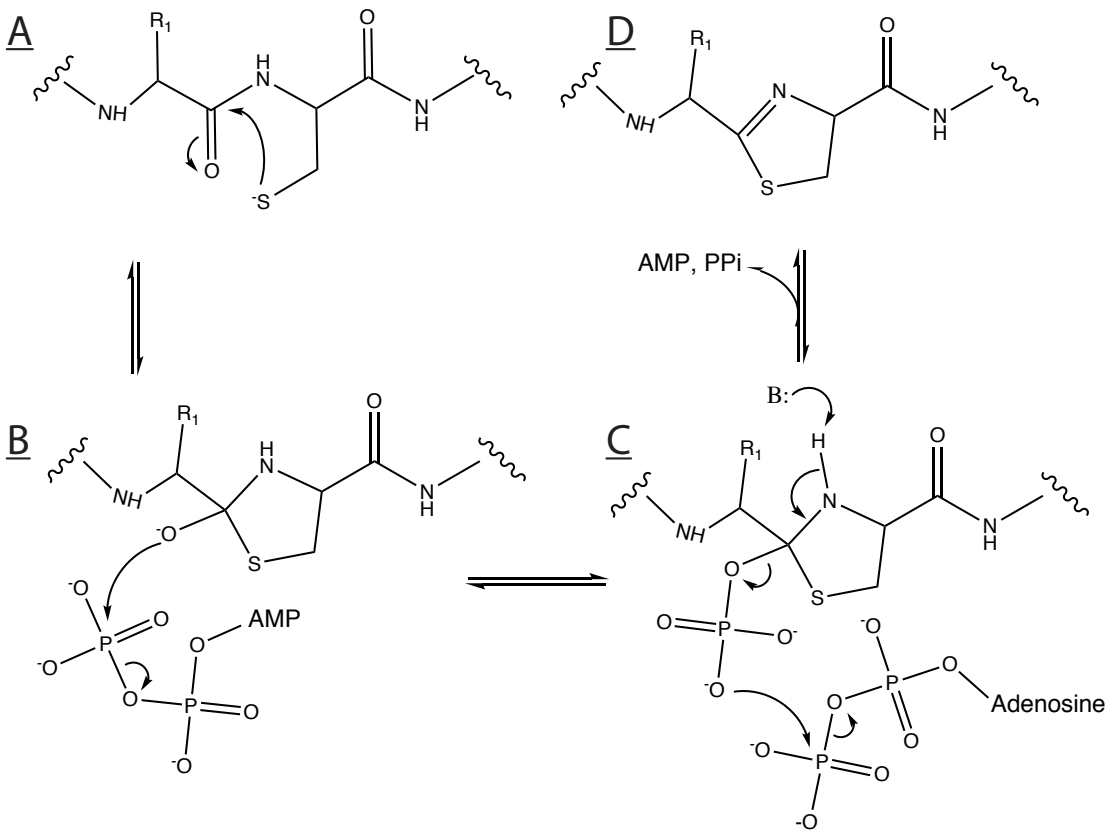

Figure 5.14: Schematic illustration of the cyclodehydration of a cysteine residue via the proposed kinase-transphosphorylation mechanism. (A) Formation of a hemiorthoamide intermediate from a cysteine residue; (B) phosphorylation of the hemiorthoamide; (C) transphosphorylation between the phosphorylated hemiorthoamide ; (D) formation of thiazoline, AMP and inorganic pyrophosphate.

detect ATP and metabolites [65]), to confirm the production of PP_i from the reaction. To further validate that PP_i is derived from both AMP-CPP and ADP, labelling of the γ-phosphate of the former and/or the β-phosphate of the latter might be required.

It is intriguing why such a process would have evolved to take place. The hydrolysis of ADP could yield energy, but formation of the unstable pyrophosphate would be energetically costly. Moreover, the transfer of a phosphate group onto the hemiorthoamide (Figure 5.14C) already introduces a good leaving group whose elimination creates the five-membered the ring structure of azolines, making it difficult to see how transphosphorylation would be preferred over simple hydrolysis. Yet if the phosphate donated by AMP-CPP can be transferred onto the β-phosphate of ADP, the phosphorylated hemiorthoamide must be stable enough to allow the release of AMP-CP and the binding of ADP. The observation of such an intermediate would provide a strong piece of evidence for this mechanisms. It is possible that cyanobactin heterocyclases are configured in a way that protects the phosphorylated intermediate, poising it for transphosphorylation, while other YcaO-family heterocyclases, which do not produce PP_i [69], are not. Although adding ADP to the AMP-CPP reaction did not accelerate heterocyclization, the duo was not a perfect mimic of ATP, owing to issues such as AMP-CP displacement. When ATP is used, the process may take place more readily. Computational analysis of the energy landscapes surrounding heterocyclization and transphosphorylation, and a structure of the ternary complex between enzyme, peptide and nucleotide might provide insights into the situation.

In summary, not only does the kinase-transphosphorylation mechanism provide a unification for the observations of ADP, AMP, P_i and PP_i made of YcaO-type heterocyclases, it has revealed a previously uncharacterised type of phosphate chemistry. Precautions should be taken, however, in ascertaining the significance of the discovery. Apart from the recommendations for further experiments stated above in this section, it would be worth checking other types of YcaO enzymes, such as the enigmatic *E. coli* YcaO protein, which is known to produce AMP and PP_i [72].

5.6 Summary of Work on Cyanobactin Enzymes

Before moving onto Chapter 6, which describes the study of an enzyme involved in histidine biosynthesis, it is perhaps pragmatic that the work laid out in this and the three preceding chapters is summarised. These chapters describe investigations into the fundamental biochemistry of several post-translational modification enzymes from cyanobactin biosynthetic pathways, and biotechnological applications based upon such research. The similarities and differences between these PTMEs are a reflection of one of the strategies used by biological systems to generate diverse secondary metabolites, that is,

ribosomal production of peptides followed by post-translational modifications. Diversity arises partly through combinations of different PTMEs, but even highly homologous enzymes, such as PatD and TruD (both heterocyclases), could carry out different modifications on the same precursor peptides. This diversity-generating functional variation in also seen in the oxidases and the prenylases, as discussed in Chapter 1. Despite our efforts, the mechanistic differences in homologous enzymes have not been deciphered on a molecular level, and may be of interest to future researchers.

On a broader scale, there are much more to be understood about PTMEs involved in RiPP biosynthesis. Not only are novel proteins and domains widely present in biosynthetic clusters, the ones with seemingly apparent functions are often found inactive. Through studies of the peptide clamp domain in the cyanobactin heterocyclase and oxidase, we show that the former is active while latter is vestigial, an observation that cautions against automatic assignment of protein functions. As commercial interests in RiPPs and the development of genome mining tools have accelerated the discovery of biosynthetic clusters through bioinformatics, an ever expanding myriad of PTMEs will provide exciting targets for academic studies and industrial exploitations alike.

Chapter 6

Structural Elucidation of ATP Phosphoribosyltransferase from *Psychrobacter arcticus*

6.1 Summary

L-Histidine is produced by bacteria, archaea, fungi and plants via a common 10-step biosynthetic pathway. ATP phosphoribosyltransferase (ATPPRT) is the first enzyme of this pathway and the key point of regulation, as it is inhibited by the end-product histidine and cellular metabolites such as AMP. ATPPRT's catalytic activity is encoded by the gene *hisG*, which exists in a long form ($HisG_L$) and a short form ($HisG_S$). The long form contains a regulatory domain in addition to the catalytic domain, which responds to histidine inhibition, while the short form has a partner protein (HisZ) that allosterically activates it and provides a different allosteric site for histidine inhibition. Although His_L is well characterised, understanding of $HisG_S$ is limited. This work describes the purification and crystallisation of the proteins HisG and HisZ from *Psychrobacter arcticus*, which form a hetero-octamer consisting of four copies of each protein.

6.2 Introduction

6.2.1 Overview of Histidine Biosynthesis

L-Histidine (histidine) is one of the 20 canonical amino acids used to build proteins in all living organisms. Its uniquely versatile role as both an acid and a base at neutral pH makes it one of the commonest active site residues. While mammals have to acquire histidine

through diet, plants, bacteria, fungi and archaea have a common histidine biosynthetic pathway that is essential to their survival [172–175]. Enzymes from this pathway therefore provide potential targets for the development of drugs against pathogenic microorganisms, such as *Mycobacterium tuberculosis* and *Salmonella typhimurium*, that exert minimal effects on their human hosts.

The histidine biosynthetic pathway involves 10 catalytic steps (Figure 6.1), starting from the compounds adenosine triphosphate (ATP) and phosphoribosyl pyrophosphate (PRPP). The reactions on the pathway are conserved across kingdoms, although genetic variations exist between organisms. ATP and PRPP are condensed by the enzyme ATP phosphoribosyltransferase (ATPPRT), product of the gene *hisG*, to generate phosphoribosyl-ATP (PR-ATP) and inorganic pyrophosphate (PP_i) [176]. PR-ATP is subsequently hydrolysed to phosphoribosyl-AMP (PR-AMP), by a PR-ATP pyrophosphatase, encoded by the gene *HisE.* and then turned into 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino) methylideneamino] imidazole-4 carboxamide (5'ProFAR) by the PR-AMP cyclohydrolase HisI, which breaks the bond between the N1 and C6 atoms of the adenine ring [177]. In organisms like *E. coli* and *S. typhimurium*, HisE and HisI are fused into a bifunctional enzyme, while in Actinobacteria such as *Corynebacterium glutamicum* they form separate polypeptides [178]. The fourth step of histidine biosynthesis is catalysed by HisA, which isomerises 5'ProFAR into 5-[(5-phospho-1-deoxyribulos-1-ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-carboxamide (PRFAR), via an internal redox reaction called Amadori rearrangement[177]. PRFAR is then split into two parts. The ATP-derived structure gives rise to 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR), which can be recycled by the *de novo* purine biosynthesis pathway. The rest of the molecule receives an amino group from glutamine to form imidazole-glycerol phosphate (IGP). This reaction is cataylzed by the IGP synthase, a heterodimer formed by the proteins HisF and HisH [177]. However, research by Kulis-Horn and colleagues suggested that *C. glutamicum* does not require the glutamine amidotransferase HisH, when growing in medium containing ammonia [178]. In the sixth step of the pathway, IGP is dehydrated by the HisB protein, and the enol intermediate is then non-enzymatically ketonised to imidazole-acetol phosphate (IAP). The oxygen is subsequently displaced with an amino group, donated by L-glutamate, yielding L-histidinol phosphate (Hol-P). Hol-P phosphatase HisN then hydrolyses the phosphate group of Hol-P. In *E. coli* and *S. typhimurium*, HisN and HisB are encoded by the same gene, *hisNB*, as a bifunctional enzyme, but in *C. glutamicum* they are encoded separately. The product, L-histidinol then undergoes two oxidations by the histidinol dehydrogenase HisD, giving rise first to L-histidinal and then L-histidine, using NADH as a co-factor.

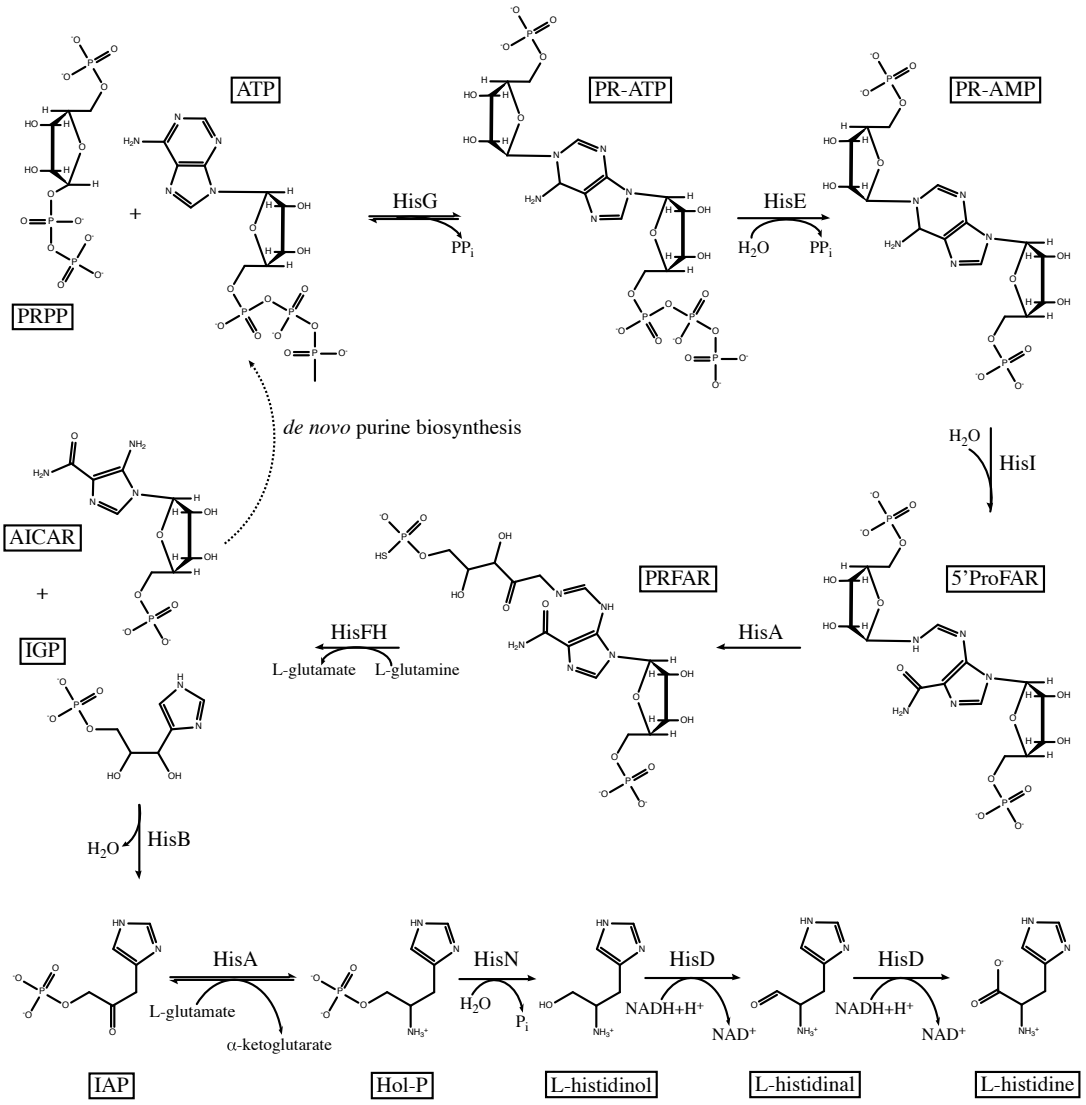

Figure 6.1: Schematic illustration of the histidine biosynthetic pathway. Substrates, products and key intermediates are highlighted in black boxes.

6.2.2 ATP Phosphoribosyltransferase

Adenosine 5-triphosphate phosphoribosyltransferase (ATPPRT, EC 2.4.2.17) is the first dedicated enzyme in histidine biosynthesis, catalysing the condensation of ATP and PRPP to generate PR-ATP and inorganic pyrophosphate (PP_i) [176]. It is found in a wide range of organisms including fungi, plants, bacteria and archaea, encoded by the gene *hisG*. ATPPRT is the key regulatory point of histidine production which responds to metabolic status of the cell/organism signalled by histidine, AMP and ADP levels [176, 179, 180]. In *Arabidopsis* the availability of ATPPRT has been shown to be the most important factor in controlling histidine levels, while that of other biosynthetic enzymes had very limited effects [181].

There are two major types of the catalytic protein HisG - the long form $HisG_L$ and the short form $HisG_S$. Biochemical properties of $HisG_L$ are well characterised [176, 179, 180, 182–192], having been subjected to investigation since mid 20th century. In addition, multiple molecular structures are available from different organisms with a range of ligands [189, 191, 193, 194]. The $HisG_L$ protomer consists of a catalytic subunit and a regulatory subunit, the latter providing a histidine binding site involved in allosteric inhibition [193]. It was originally thought that the enzyme is active in its homodimeric form, but forms an inactive homohexamer when the concentration of either the enzyme or histidine was high [193]. Later reports [188–190] however showed that both the active and inactive enzyme forms are predominantly hexamers in solution, with ligand-induced conformational changes, rather than variations in quaternary structure, being responsible for inhibition. More recently, crystal structures of the *Medicago truncatula* ATPPRT ($HisG_L$) captured the enzyme in a relaxed and a tense state, corresponding to the free (active) and the histidine-bound forms, respectively [195].

On the other hand, relatively little was known about $HisG_S$. Lacking the the regulatory domain, $HisG_S$ does not bind histidine, but instead forms a equimolar hetero-octameric complex with a partner protein, HisZ, which activates $HisG_S$ while simultaneously rendering it susceptible to histidine inhibition. HisZ is a paralogue of histidyl-tRNA synthetase [196, 197], although no activity of HisZ itself has been reported. Prior to this study, it was believed that $HisG_S$ is active only as the hetero-ocatmeric holoenzyme [196–198]. Now it has been demonstrated that $HisG_S$ is also active as a homodimer, albeit showing reduced efficiency [199, 200]. Due to the insensitivity to histidine inhibition brought by the absence of HisZ, the homodimeric form of the enzyme displays increased potentials as a biocatalyst compared to the hetero-octameric holoenzyme and $HisG_L$. Two crystal structures of short form ATPPRTs were available (Figure 6.2), both hetero-octamers, from *Lactococcus lactis* [196] and *Thermotoga maritima* [197]. In both structures, HisG and HisZ are organised in dimers. Two dimers of HisZ pair up in an crossed-over manner and form the core of the octamer, while one pair each of HisG dimer

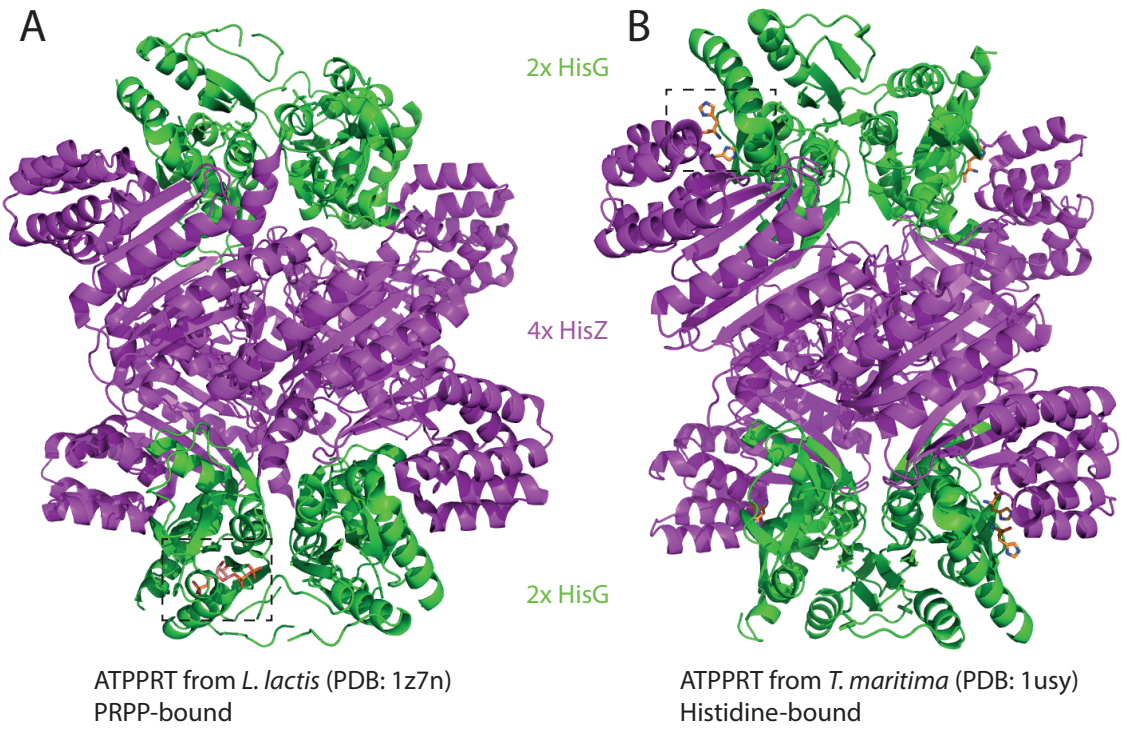

Figure 6.2: Structure of ATPPRT from (A) *Lactococcus lactis* (PRPP-bound, PDB: 17ZN) and (B) *Thermotoga maritima* (histidine-bound PDB: 1USY). HisG and HisZ subunits are shown in green and magenta, respectively. Black dashed boxes highlight the bound ligands.

occupy on the top and bottom of this core.

A recent study showed that both forms of ATPPRT operate via the same transition state [192], establishing the relatedness of $HisG_S$ and $HisG_L$, despite their different modes of regulation.

6.2.3 Aims and Objectives

The work described in this chapter was carried out in order to determine the molecular structure of ATPPRT from *Psychrobacter arcticus*, a cold adapted enzyme that was first isolated from Siberian permafrost. The *hisG* gene in this organism encodes a short form HisG, and a *hisZ* gene is present. As the two existing structures of $HisG_S$-type ATPPRT are from mesophilic [196] and thermophilic [197] organisms, a cold-adapted version of this enzyme would be a valuable addition to the understanding of the family. This project was carried out in collaboration with Dr Rafael da Silva's group, the members of which characterised *Pa*ATPPRT kinetically.

6.3 Materials and Methods

6.3.1 Expression and Purification of *Pa*HisG and *Pa*HisZ

Condon-optimised genes encoding *P. arcticus* HisG and HisZ were purchased from DNA 2.0 on a pJexpress431 and a pJexpress414 vector, respectively. The plasmid for *Pa*HisG was transformed into *E. coli* C43 (DE3) cells, and recovered from an LB-kanamycin agar plate. A single colony was inoculated into 100 ml LB broth supplemented with 50 µg/ µl kanamycin and incubated at 37 °C for 16 h, before 10 ml of which was transferred to 1 litre of fresh LB-kanamycin medium. The culture was incubated at 37 °C until its OD_{600} reached 0.8, at which pointed the temperature was lowered to 16 °C . ITPG (to the final concentration of 1 mM) was added to the culture after 30 min of equilibration at 16 °C, and the cells were allowed to overexpress the protein for 20 h. Cells were harvested by centrifugation at 8983 *g* for 15 min and stored at -80 °C. The expression construct for *Pa*HisZ was transformed into *E. coli* BL21 (DE3) cells, and a single colony recovered from LB-kanamycin agar plates was inoculated into 100 ml LB broth supplemented with 100 µg/ µl ampicillin. After growing at 37 °C for 16 h, 10 ml of the seed culture was transferred to 1 L of fresh LB-ampicillin medium. The incubation, expression and harvest methods were identical to those used for *Pa*HisG. Both constructs contained an N-terminal cleaveable TEV protease site.

*Pa*HisG and *Pa*HisZ were purified separately and all steps were carried out at 4 °C or on ice. Cells were resuspended in HisG/Z lysis buffer (50 mM HEPES, pH 8.0, 10 mM imidazole and 500 mM NaCl), supplemented with 0.2 mg/ml lysozyme, 0.4 mg per g of cell DNase and 1 tablet of complete EDTA-free protease inhibitor (Roche) per 50 ml of suspension. The lysate was passed through a cell disruptor (Constant Systems) at 30 kpsi and clarified by centrifugation at 48384 *g* for 20 min at 4 °C, followed by filtration through a Millipore nitrocellulose membrane with a pore size of 0.45 µm (Sigma). Filtered lysate was applied to a Ni-NTA column (GE Healthcare), equilibrated with lysis buffer, and eluted with a gradient of 0-60% HisZ/G elution buffer (50 mM HEPES, pH 8.0, 500 mM imidazole and 500 mM NaCl) over 20 min. Elution fractions were subjected to SDS-PAGE analysis and those that contained the protein of interested were collected and dialysed into HisG/Z dialysis buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 2 mM DTT, 10% Glycerol). TEV protease (1:10 mass ratio) was added for the removal of his-tag. The cleavage processed was carried out at 4 °C for 6 days, with one changes of buffer in between and finally re-dialysed into HisG/Z lysis buffer. The portein was then passed through a second Ni-NTA column equilibrated with HisG/Z lysis buffer. The flow-through was collected, concentrated and dialysed into 20 mM HEPES, pH 8.0, before being flash frozen and stored at -80 °C. The identity and purity of the proteins were confirmed by SDS-PAGE and mass spectrometry.

6.3.2 Crystallisation of *Pa*ATPPRT

Purified *Pa*HisG and *Pa*HisZ were mixed at a 1:1 molar ratio and subjected to PCT pre-crystallisation test (Hampton Research) to determine the optimum concentration for crystallisation. The protein mix was concentrated to 8 mg/ml and mixed with various concentrations of ATP, NaCl, $MgCl_2$ and or L-histidine, before being subjected to stochastic screening in 96-well Intelli-Plates Gryphon crystallization robot (Art Robbins Instruments), with protein:precipitant mixed at 1:1 and 2:1 volume ratios. Hits were visualised on a Rigaku Minstrel HT UV instrument. The protein buffer used which yielded diffracting crystals was 20 mM Tris-HCl, pH 7.0 supplemented with 50 mM KCl, 10 mM ATP, 10 mM $MgCl_2$, 2 mM DTT and 0.5 mM L-histidine. Initial hit was obtained in 0.1 M Bicine pH 8.5, 15.73% PEG 3350, 0.12 M $CaCl_2$, and 2.16% 1,6-hexanediol. Optimisation was carried out using the hanging-drop vapour-diffusion method in 24-well plates with protein and precipitant mixed in a 1:1, 1:2 and 2:1 volume ratios. Crystals (0.3 mm x 0.1 mm x 0.1 mm) were obtained from the optimised condition (0.1 M Bicine pH 8.5, 11% PEG 3350, 0.15 M SrCl2, 0.15 M KBr, and 2% 1,6-hexanediol) after growing for 21 days at 4 °C.

6.3.3 X-ray Data Collection and Processing

Crystals were cross-linked by vapour diffusion with a 25% (v/v) aqueous solution of glutaraldehyde for 1 h prior to being cryoprotected in 15% 2,4-methylpentanediol (MPD) freshly prepared in a reservoir solution and then flash-cooled in liquid nitrogen. Diffraction data were collected on a Rigaku 007 MM HFM generator with a Saturn 944+ CCD detector. Data were processed with iMosflm [201] and scaled with Aimless [202].Matthews coefficient [203] calculations indicated that the asymmetric unit contained four copies each of *Pa*HisG and *Pa*HisZ. The structure was determined using molecular replacement (with the help of Dr Garib Murshudov, University of Cambridge) employing the Phaser programme [204] from the CCP4 Suite [205]. The following 3 structures were used as search models: Protein Data Bank (PDB) entries 1Z7M (*L. lactis* ATPPRT, containing both subunits) [199], 2VD2 (HisG homologue from *Bacillus subtilis* , unpublished), and 3OD1 (HisZ homologue from *Bacillus halodurans*, unpublished). CCP4 Buccaneer [206] was used to assign the majority of correct side chains to the *Pa*ATPPRT model. Missing side chains and residues were added with Coot [207] and the model was refined with Refmac [208].

6.3.4 In-solution Cross-Linking of *Pa*HisG and *Pa*HisZ

*Pa*HisG (40 µM), *Pa*HisZ (40 µM) and an equimolar mixture of the two proteins (40 µM each), in the buffer 100 mM HEPES pH 8.0, 100 mM KCl, 10 mM MgCl2 and 4 mM DTT, were applied to a coverslip and sealed in a 24-well crystallisation plate. A solution of 25% glutaraldehyde was also applied to the coverslip and used to cross-link to proteins via vapour diffusion at 4 °C for 16 h. The process was repeated with 2 mM ATP, 1 mM histidine or both added to the protein buffer, and with elevated protein concentrations (133 µM, concentration used crystallisation). The cross-linked proteins, along with an uncrosslinked 1:1 mixture of *Pa*HisG and *Pa*HisZ (40 µM or 133 µM each), were analysed by SDS-PAGE on a NuPAGE Novex 4-12% Bis-Tris gel (Invitrogen) against the Precision Plus Protein Dual Color Standards (Bio-Rad) marker.

6.3.5 Multiangle Light Scattering Analysis of *Pa*HisG

To determine the oligomeric status of *Pa*HisG in solution, purified *Pa*HisG (3.3 mg) was subjected to size exclusion chromatography (SEC) on a HiLoad 16/600 Superdex 200 pg column (GE healthcare) equilibrated in 20 mM HEPES pH 8.0. The sample was eluted in the same buffer at a flow rate of 1 ml/min. The flow was passed through and analyzed by multi-angle light scattering (MALS) detector DAWN HELEOS-II and refractive index detector Optilab T-rEX (Wyatt Techonologies). The elution peak, shown as changes in refractive index, was divided into 1 mL fractions in the ASTRA 6.1 software (Wyatt Techonologies) for calculations of molecule size in mean square radius (nm^2) and molecular mass.

6.4 Results

6.4.1 Purification of *Pa*HisG and *Pa*HisZ

*Pa*HisG (25.2 kDa) and *Pa*HisZ (43.1 kDa) were purified to homogeneity by Nickel affinity chromatography, followed by TEV-cleavage of the N-terminal His-tag and a second Nickel column. The purified samples were subjected to SDS-PAGE (Figure 6.3) and mass spectrometry to confirm their purity and identity. Size-exclusion chromatography was not performed as both subunits were assumed to be in an equilibrium of different oligomeric status, which will change upon mixing.

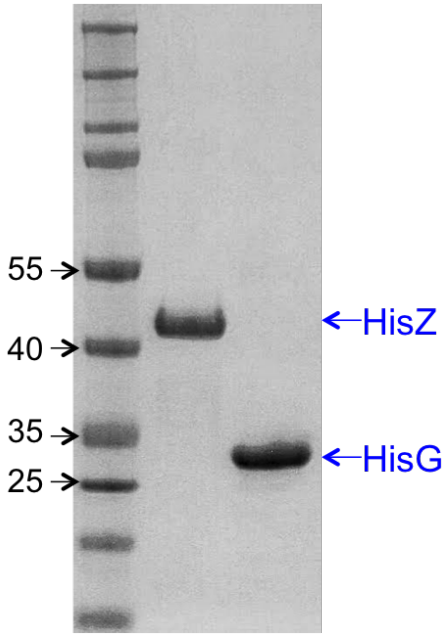

Figure 6.3: SDS-PAGE analysis of *Pa*HisG and *Pa*HisZ.

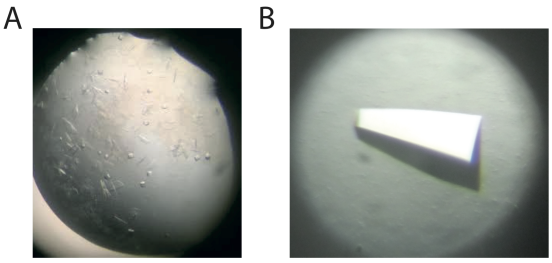

Figure 6.4: Crystallisation of *Pa*ATPPRT. (A) Crystals from the initial hit condition; (B) a crystal from the optimised condition (size 0.3 mm x 0.1 mm x 0.1 mm).

6.4.2 Crystallisation of *Pa*ATPPRT

A 1: 1 molar mix of *Pa*HisG and *Pa*HisZ (8 mg/ml) was subjected to stochastic screens, and initial crystals (Figure 6.4A) were found in the condition 0.1 M Bicine pH 8.5, 15.73% PEG 3350, 0.12 M $CaCl_2$, and 2.16% 1,6-hexanediol. These crystals exhibited two different morphologies, a relatively long but flat one, and a hexagonal cylindrical one. The crystallisation condition was optimised to 0.1 M Bicine pH 8.5, 11% PEG 3350, 0.15 M $SrCl_2$, 0.15 M KBr, and 2% 1,6-hexanediol, in which larger crystals formed (Figure 6.4B). Crystals grown in this condition were of the elongated form, but has increased in thickness. The crystal shown in Figure 6.4B was 0.3 mm x 0.1 mm x 0.1 mm in size.

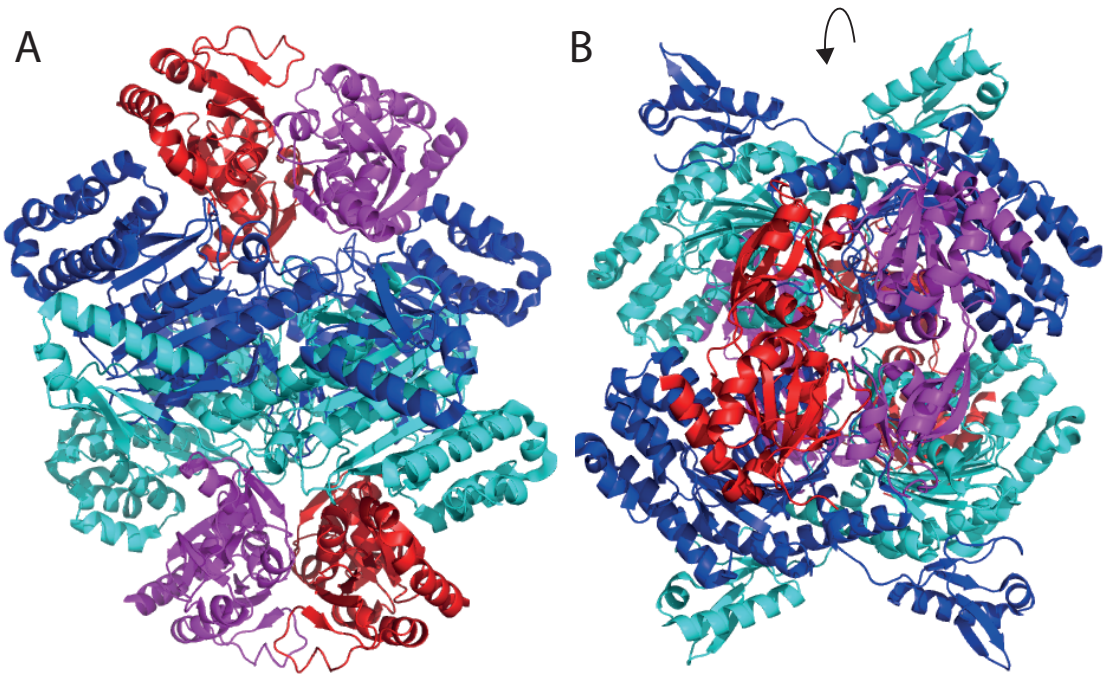

Figure 6.5: Structure of ATPPRT from *Psychrobacter arcticus* (PDB: 5M8H). HisG subunits are coloured in red and magenta; HisZ subunits are coloured in blue and cyan. (A) front view; (B) top view.

6.4.3 Crystal Structure of *Pa*ATPPRT

The three-dimensional structure of the *Pa*ATPPRT was determined and refined to 2.34 Å resolution. Refinement statistics are summarised in Table 6.1, and atomic coordinates were deposited in the PDB as entry 5M8H. Four copies each of HisG and HisZ formed a hetero-octamer, arranged in the same architecture as the *L. lactis* and the *T. maritima* ATPPRTs (Figure 6.2). Two HisZ dimers form the middle section of the complex, while two dimers of HisG cover the top and bottom of the structure (Figure 6.5). Although 10 mM ATP and 0.5 mM histidine were present in the crystallization mixture, no clear electron density was observed for either ligand. Density for glutaraldehyde was not also not observed, suggesting that the proteins may not have been cross-linked in the crystal.

6.4.4 HisG Subunits

The four HisG molecules in this complex are homologous and overlay with an r.m.s.d. of 0.32, 0.18, and 0.38 Å over 208 Cα atoms, when comparing chain E to chains F, G and H, respectively. PaHisG displays two domains (Figure 6.6A). Domain 1 comprises residues

Table 6.1: Crystallographic data and refinement statistics of *Pa*ATPPRT. Numbers in brackets describe the highest-resolution shell. $R_{merge} = (\sum(I_{hkl} - I) \times 100)/\sum I_{hkl}$ (I = average intensity, I_{hkl} = measured intensity for a given observation); $R_{work} = \sum_{hkl}(F_o hkl - F_c hkl)/\sum_{hkl} F_o hkl)$, with $F_o hkl$ and $F_c hkl$ being the observed and calculated structure factor amplitudes, respectively; $R_{free} = R_{factor}$ for a test set of reflections (5%). Table taken from [200].

PDB entry	5M8H
wavelength (Å)	1.5418
resolution (Å)	30.63-2.34 (2.38-2.34)
total no. of reflections	390223
no. of unique reflections	110587
multiplicity	3.5 (2.4)
completeness (%)	97.6 (81.0)
I/σI	6.3 (0.9)
R_{merge}	0.135 (0.969)
CC1/2	0.994 (0.366)
Wilson B (Å^2)	34.1
space group	$P2_1$
unit cell parameters	
a (Å)	94.02
b (Å)	146.73
c (Å)	101.90
α (deg)	90.00
β (deg)	102.4
γ (deg)	90.00
Refinement	
no. of reflections (working set)	105092
no. of reflections (R_{free} test set)	5467
R_{work}, R_{free}	0.23, 0.27
rmsd for bond distances (Å)	0.010
rmsd for bond angles (deg)	1.438
no. of molecules per asymmetric unit	4 HisG , 4 HisZ
no. of protein atoms	18229
no. of ligands	37
no. of water molecules	629
average B factor (Å^2)	
protein	50
ligands	58
waters	36
Ramachandran plot	
favored (%)	97.4
allowed (%)	2.6
outliers	0

1-115 and 197-228, whilst residues 115-197 (residues preceding E20 are disordered and have not been included in this model) form the second domain (domain 2). Domain 1 has 6 β-stranded core surrounded by 5 α-helices. The central β1 forms a 3-stranded parallel sheet with β3 and β4, each of which in turn form 2 small anti-parallel sheets in conjunction with the lateral strands (β2-β3; and β4-β9-β5). These sheets are encircled by helices α1, α2, α3, α7, and α8. Domain 2 displays a similar global fold, with 5 β-strands arranged as a 3-stranded parallel sheet (β7-β6-β8) and a 3-stranded antiparallel sheet (β8-β5-β9). The strands are sandwiched between α4 on one side and α5 and α6 on the other. β5 and β9 are common to both domains. Overlaying domain 2 on to domain 1 results in an r.m.s.d. of 3.5 Å over 95 Cα atoms, despite the architectural similarities between the two.

Each of the two homodimers of HisG in the HisGZ complex comprises a head to tail arrangement (Figure 6.6B), with domain 1 of monomer 1 interacting with domain 2 of monomer 2, and vice versa. The PISA server [210] gave a CSS score of 0.343 between subunits E and H (CSS scores vary from 0, which indicates no interaction, to 1, which means the interactions are essential to complex formation). The interactions cover approximately 1190 Å^2 of surface from each monomer, equivalent to 10.8% of the solvent-accessible area. 31 residues (Figure 6.6C) contribute to the interface, mainly from α2, β2, β3, β7, and α5. R73 (chain E only), D76, Y159, and G170 formed a hydrogen bond network (Figure 6.6F), while hydrophobic residues including L58, I59, L70, L72, Y80, A85 from domain 1, and L164, V168, and L170 from domain 2 were buried by the interface.

Residues 163-184 of *Pa*HisG forms a conserved ATPPRT signature motif (identified by PROSITE [211]) that contains the phosphate-binding P-loop (residues 179-183) - a commonly found motif in ATP and GTP-binding proteins. Although the electron density observed was insufficient for substrates to be modelled in the structure, the location of the active site can be inferred from that identified in homologues, to lie in a crevice between domain 1 and domain 2 (Figure 6.6E). Side chains that interact with PRPP in the *Ll*ATPPRT structure [196], are conserved in *Pa*HisG (Figure 6.6F), namely E142, D155, E158, T159 and T162 (E163, D176, D179, T180, T183 of *Pa*HisG).

6.4.5 HisZ subunits

The four HisZ molecules overlay with an r.m.s.d. of 0.80, 0.92, and 0.13 Å (chains B, C and D, respectively, compared to chain A). *Pa*HisZ comprises two distinct domains (Figure 6.7A), an N-terminal domain (residues 1-318) and a C-terminal domain (residues 324-387), connected by an elongated loop. The four HisZ molecules are arranged in two homodimers (Figure 6.7B) in the centre of the hetero-octameric complex. As well as an extensive N-terminal interface, the C-terminal domain from each HisZ molecule crosses over the dimeric interface to form further interactions with NTD of the other monomer.

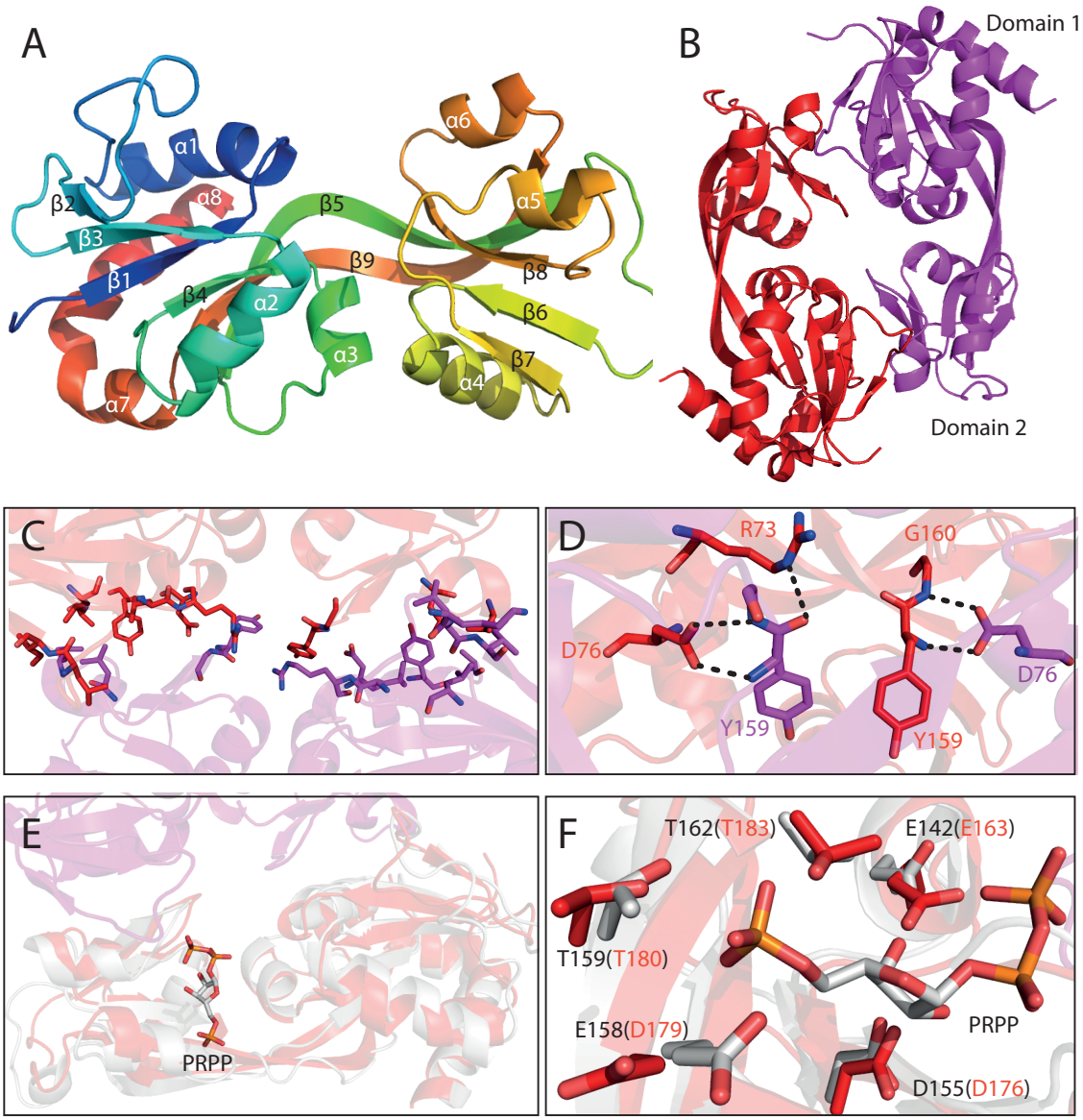

Figure 6.6: The *Pa*HisG Subunit. (A) Secondary structures in *Pa*HisG assigned by STRIDE [209]. α1-α8 designates α-helices while β1-β9 denotes β-strands. The molecule is coloured on a reverse rainbow spectrum from N- to C-terminus. Labels are coloured black or white for illustrative clarity. (B) Dimer of *Pa*HisG shown as chain E (red) and chain H (purple) of PDB entry 5M8H. (C) Residues involved in the homodimeric interface, as identified by the PISA server [210]. (D) Residues participating in hydrogen bonding at the interface. Black dotted lines represent hydrogen bonds. (E) Alignment with *Ll*HisG (white, PDB: 1Z7N, chain F), showing the location of the active site. (F) Active site residues in *Ll*HisG (residues in white sticks and labels in black) and *Pa*HisG (residues in red sticks and labels in red).

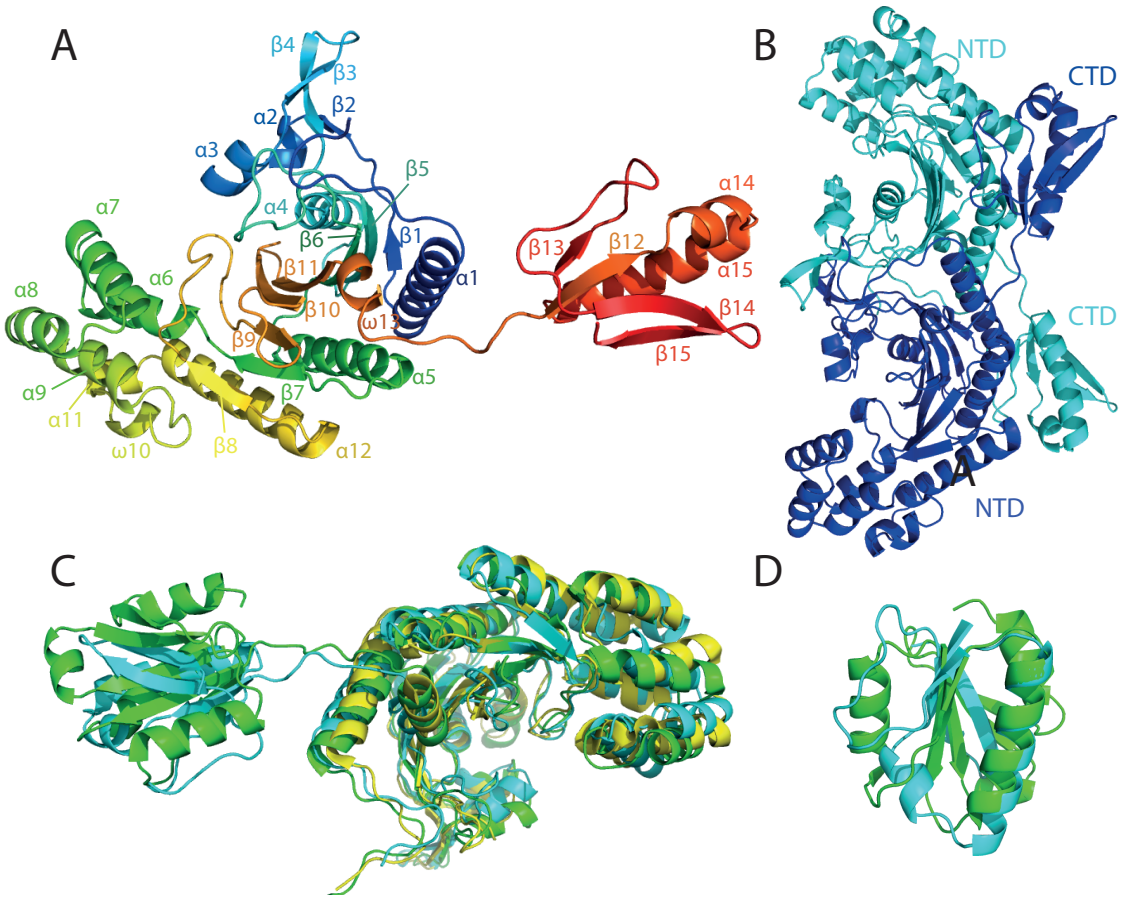

Figure 6.7: The *Pa*HisZ Subunit. (A) Secondary structures in *Pa*HisZ assigned by STRIDE [209]. α, β and ω denote α-helices, β-strands and 3-10 helices, respectively. The molecule is coloured on a reverse rainbow spectrum from N- to C-terminus. Labels are coloured as the secondary structure they represent. (B)The HisZ homodimer between chain A (blue) and chain C (cyan). (C) Overlay of *Pa*HisZ (cyan), *Ll*HisZ (1Z7N, yellow) and human HisRS (1WU7, green), showing that *Ll*HisZ lacks the CTD that is common to the other two. (D) Overlay of CTDs of *Pa*HisZ (cyan) and human HisRS (green).

When viewing the *Pa*ATPPRT as a whole, C-terminal domains of HisZ face outwards (Figure 6.5B), and have few interactions with other parts of the enzyme.

In each HisZ subunit, a six-stranded anti-parallel β-sheet (β5-β6-β11-β10-β9-β7) core, flanked by two smaller strands (β1, β8) parallel to either end, forms the centre of the N-terminal domain, with β1 on the homodimeric interface side and β8 near the periphery of the complex. The sheets are embraced by eight alpha or 3-10 (denoted with 'ω') helices (α1, α5-9, ω10, ω11 and α12), which occupy the peripheral side of the subunit, forming a bowl-like structure. Three long alpha helices, (α1, α5 and α12) runs along the bottom (as in Figure 6.7A) of the beta sheets, while the smaller helices (α6-9, ω10 and ω11) stacks on top of α12 in two columns, forming an L-shaped corner and extending to the other side of the β-sheet core. On the other side of the beta sheet core is another alpha helix (α4) and above it a 2-helices (α2, α3) and 3-strands (β2-4) insertion. The C-terminal domain

comprises 4 β-sheets (β12-15), and 2 α-helices (α14-15).

PISA server gives the HisZ homodimer a relatively low interface score of 0.231 (chains C and A), despite the apparently entangled arrangement of the two monomers and the involvement of both the CTD and the NTD. The interface covers 3313.7 Å^2, which accounts for 15.6% of the solvent-accessible area. A total of 30 hydrogen bonds and 27 salt bridges are involved in homodimer interactions. Interface residues are clustered in both the N- and C- terminal areas, mainly in the regions 1-134 and 352-369, with only one exception (N248).

HisZ has shown evolutionary links with the aminoacyl(histidyl)-tRNA synthetase HisRS [212]. Interestingly, nine out of the ten top hits from the DALI server (identifies structural homologues) [213] are HisRSs. The structure of *Pa*HisZ resembles several of these protein, such as 3OD1 from *Thermoplasma acidopilumand* and 1WU7 from Homo Sapiens, more than it does *L. lactis* (1Z7M) *T. maritima* (1USY) HisZs. The main difference between *Ll*HisZ/*Tm*HisZ and *Pa*HisZ is that *Pa*HisZ is longer than the other two, with the additional residues corresponding to its CTD, which is absent in the other two HisZs that have been structurally characeterised (Figure 6.7C). The CTD of *Pa*HisZ overlays with that of *Hs*HisRS with an r.m.s.d of 0.87 over 93 Cα atoms (Figure 6.7D). It is unclear what function does C-terminal domain serve in *Pa*HisZ; most likely it is an evolutionary remnant from a common ancestor with HisRS.

6.4.6 HisG-HisZ Interface

Each of the four HisZ subunits form contacts with two HisG subunits that are homodimeric partners with one another (Figure 6.8A). The NTD of HisZ interacts with domain 1 of one HisG and domain 2 of the other. Neither group of interactions were very extensive, with PISA giving a CCS store of 0.025 to the first group and 0.012 to the second. The total interface area between HisZ and domain 1 and 2 of HisG are around 850 Å^2 and 350 Å^2, respectively. Multiple hydrogen bonds and salt bridges were found at both interfaces (Figure 6.8B), involving residues K93, N102, Y105, E106, Q206, K210, K244 of HisG and D155, A184, N185, K186, L188, E190 and D256 of HisZ at one site, and residues P3, D12 (HisZ) and N129, Q152 (HisG) at the other. Additionally, multiple hydrophobic residues, primarily from HisG, are buried in the surface, contributing to the binding between the subunits.

6.4.7 B-Factor Analysis of *Pa*ATPPRT

Enzymes from psychrophilic organisms must adapt to maintain catalytic activity within cold environments, as chemical reaction rates decrease with temperature; for a mesophilic

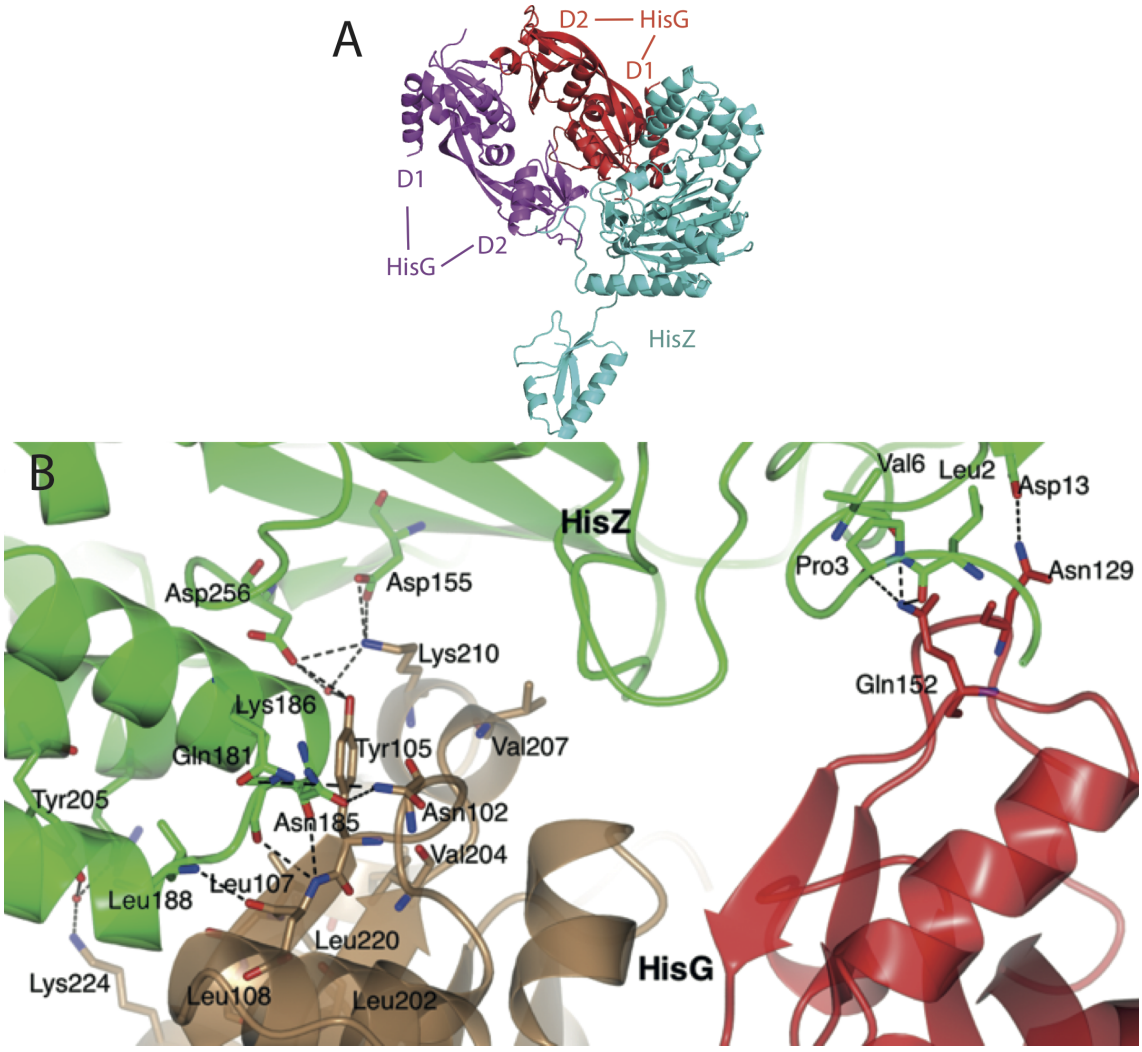

Figure 6.8: Interface Between *Pa*HisZ and *Pa*HisG Subunits. (A) Interfaces are found between HisZ (cyan) and domain 1 of one HisG subunit (red) and domain 2 of its partner (magenta). (B) Interface between HisZ (green) and HisG (red and brown) subunits showing interactions between residues. Black dotted lines indicate polar contacts. (B) was created by Dr Magnus Alphey.

enzyme, the impairment on rate can be up to 250-fold when the temperature drops from 37 °C to 0 °C [214]. Reactions catalysed by cold-adapted enzymes tend to have a lower activation enthalpy and a more negative entropy [215–217], which reduces the effect of temperature on their reaction rates, but the structural basis of these thermodynamic adaptations are not fully understood. It is thought that global or local (active site) flexibility could contribute to increase activity, and the choice between these two strategies may lie in the size of the substrate, with enzymes using larger substrates exhibiting global flexibility, and vice versa [214, 218]. An alternative, emerging theory based on computer simulations states that increased flexibility at protein surfaces, rather than the active site, is key to cold adaptation [219, 220].

The distribution of B-factors throughout the *Pa*ATPPRT holoenzyme was analysed on shown in Figure 6.9. Overall the hetero-octamer appears to be 'hotter' on the outer side Figure 6.9A. The inflexible regions are concentrated in the HisZ subunits, which situate in the middle of the protein. Within a HisZ subunit (Figure 6.9C), it is 'colder' in the core and normal on the peripherals. The catalytic subunit, HisG, (Figure 6.9B) is more flexible. However, within this subunit the outer regions again appear to be 'hotter' than the active-site cleft (shown by the thickness of the 'tube'). As both the catalytic site and the periphery of the protein are more flexible compared to the core, it is not obvious how much each factor contributed to the psycrophilicity of the enzyme. Computational analysis of *Pa*ATPPRT and its mesophilic/thermophilic homologues might provide insight on the thermodynamics of these enzymes.

6.4.8 Oligomeric Status of *Pa*HisG and *Pa*HisZ in Solution

To confirm that the hetero-octameric form is the active enzyme in solution, rather than a crystallographic artefact. cross-linking experiments with glutaraldehyde were carried out. It was presumed that if two or more subunits are in close proximity with each other in solution, glutaraldehyde (covalently links lysines) will introduce covalent bonds between them that are resistant to boiling and electrophoresis. HisG, HisZ or a mixture of both (40 µM each - at this concentration both HisG and the holoenzyme have exhibited enzymatic activity [200]) were treated with glutaraldehyde via vapour diffusion and then analysed via SDS-PAGE (Figure 6.10). The experiments were repeated with 2 mM ATP, 1 mM histidine, or both added to the protein to see if ligand-binding changes the oligomeric state of the subunits. For both HisG and HisZ (Figure 6.10, left panel), the monomeric and homodimeric forms are the most abundant after cross-linking, while higher oligomers were also observed. The addition of ATP and/or histidine did not alter the patterns (middle panel). When HisG and HisZ were mixed, new bands appeared (left panel, lanes 7 and 8), corresponding to the formation of hetero-oligomers. Several of the newly observed bands were subjected to mass spectrometry (tryptic digest), and both HisG and HisZ

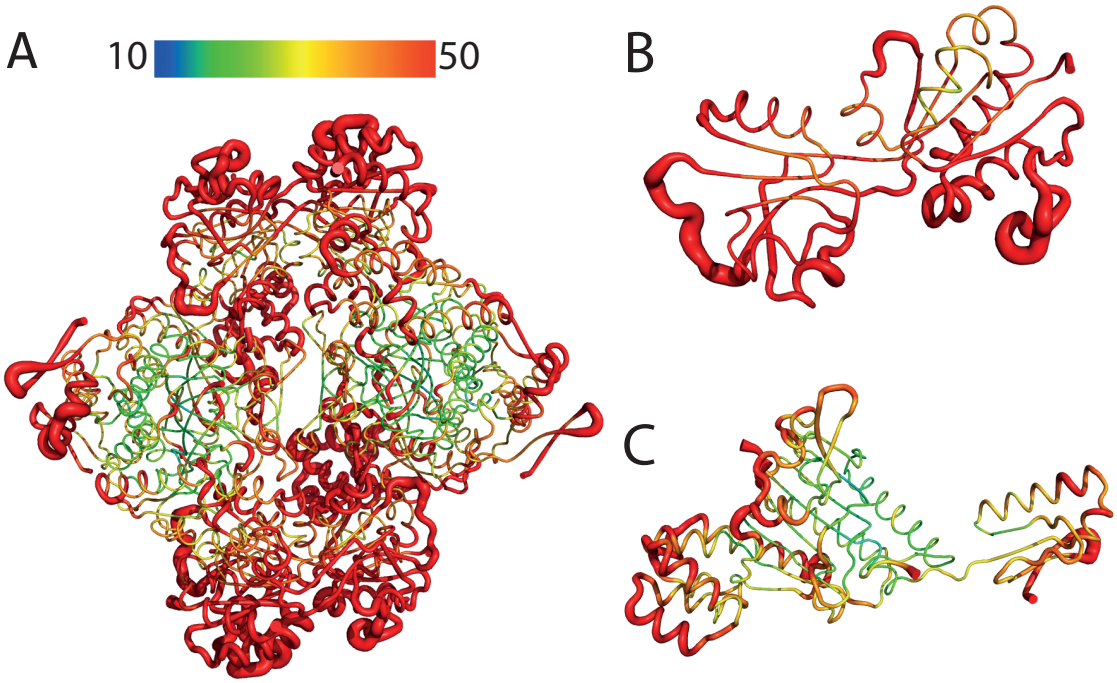

Figure 6.9: B-factor distribution in *Pa*ATPPRT. (A) B-factor distribution throughout the *Pa*ATPPRT holoenzyme. The protein is coloured based on b-factors on a rainbow spectrum with higher b factors shown in warmer temperatures. The thickness of the 'tube' also increases with b-factor values.(B) B-factor distribution in the HisG subunit. (C) B-factor distribution in the HisZ subunit.

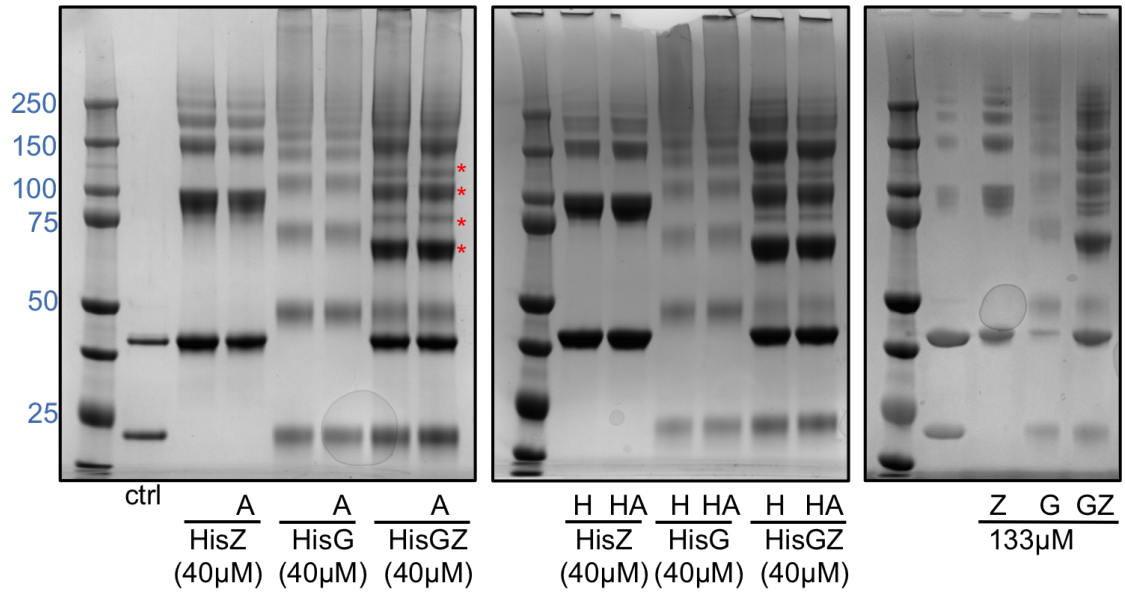

Figure 6.10: SDS-PAGE of Cross-Linked *Pa*HisG and *Pa*HisZ. Ctrl, uncross-linked HisG and HisZ (mix); A, 2 mM ATP; H, 1 mM histidine; G, HisG; Z, HisZ. Bands marked with red asterisk were subjected to MS analysis.

were identified in all of them. The most intense band located between the 50 kDa and 75 kDa markers, was mostly likely a heterodimer (calculated MW = 68.3 kDa), while the identity of higher bands were more difficult to interpret. Regions above the 250 kDa marker were smeared and it was not possible to see a clear band for the hetero-octamer (273 kDa). Raising the concentration of proteins to 133 µM to match that used in the crystallisation trials yielded highly similar results (Figure 6.10, right panel).

While the results were different to what we expected and the hetero-octameric form was not clearly identifiable, the consistency between lower (40 µM) and higher (133 µM) enzyme concentrations means that oligomeric states are unlikely to differ significantly between the two. Later kinetic and structural studies [200, 221, 222] have shown that allosteric activation requires that HisZ binds to HisG in the hetero-octameric fashion. The reason for the tendency towards lower oligomers may have been due to the ineffectiveness of cross-linking by glutaraldehyde. Lysine residues are found at both homodimeric (such as K61 from HisZ and K157 of HisG) and heterodimeric (such as K186 from HisZ and K210 from HisG) interfaces in the crystal structure, but they were not seen cross-linked, even though the crystals were treated with glutaraldehyde. On the other hand, free surface lysine residues that are not involved in interface formations (e.g. K242 of HisZ) could potentially be cross-linked to other non-interface residues when diffusion of molecules placed them in the vicinity of each other, forming complexes that showed on the SDS-PAGE analysis but lacked physiological relevance.

Due to the observation that *Pa*HisG alone exhibits ATPPRT activity [200], the oligomeric status of HisG in solution was also investigated. Purified protein was subjected

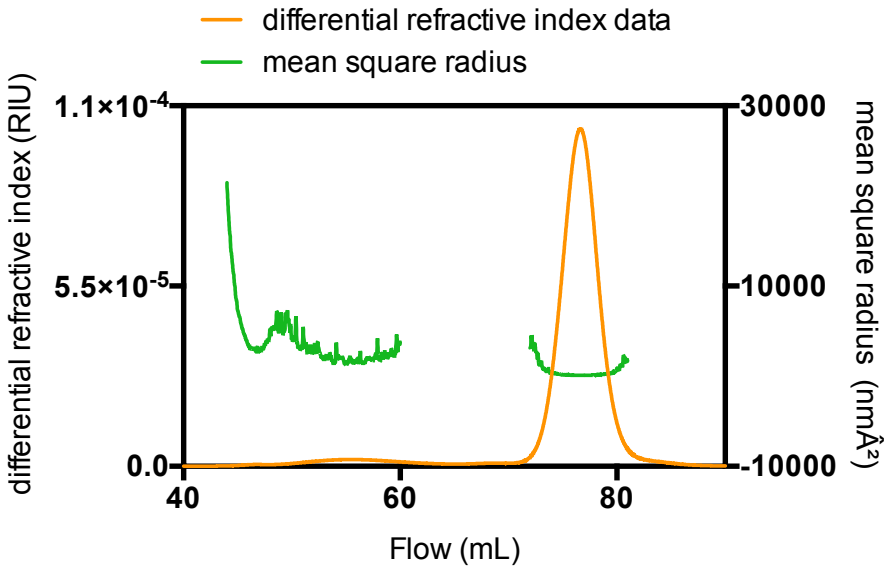

Figure 6.11: SEC-MALS analysis of *Pa*HisG. Refractive index changes when a substance that differs from the buffer (in this case, the protein) passes through the detector; Mean square radius is correlated to molecular weight.

to SEC-MALS and eluted primarily as a highly homogenous peak (Figure 6.11), with a calculated molecular mass of 47.5 kDa, close to the predicted mass of a homodimer (50.4 KDa). A small amount of higher oligomers were also present.

6.5 Discussion

The structure of *Pa*ATPPRT added to the existing family of enzymes that employ a short-form HisG and a regulatory subunit HisZ. Although the structure described here did not show any ligands, further work led by Alphey *et al.* [221] captured this enzyme in complex with PRPP, PRPP+ATP, PRPP+ADP and PR-ATP. In the meanwhile, kinetic analysis by Fisher *et al.* [222] also established that PRPP forms a binary complex with the enzyme before ATP is able to bind. The order of binding may explain the inability to achieve complex structures in this study.

Efforts were also made to crystallise the enzyme in complex with histidine. The *Tm*ATPPRT structure was histidine-bound due to the presence of histidine in their crystallisation buffer. However, co-crystallisation and soaking were both attempted for *Pa*ATPPRT, with no success; oftentimes the crystals were visibly damaged or their diffraction quality was significantly poorer (6 Å or worse) compared to unsoaked ones (around 3 Å). These experiments were mainly carried out with histidine as the sole ligand or with ATP+histidine, and on a few occasions with ATP+PRPP+histidine. It is possible that histidine binds the PRPP-enzyme binary complex or the product (PR-ATP and P_i)-

enzyme ternary complex. It could be worthwhile to determine the mode of inhibition by histidine kinetically in addition to crystallisation experiments.

The finding that *Pa*HisG is primarily homodimeric in solution is consistent with later studies that crystallised it as a dimer [221]. In addition to *Pa*HisG, *Ll*HisG$_S$ [199] and a truncated form of *Campylobacter jejuni* HisG$_L$ containing only the catalytic domains [223] also demonstrated ATPPRT activities. Due to the absence of histidine inhibition, these enzymes could potentially be exploited as biocatalysts for histidine synthesis. *Pa*HisG may have a particular advantage in this aspect as psychrophilic enzymes tend to have higher activities at lower temperatures, which could result in saving a sizeable amount of energy when employed on an industrial scale.

Bibliography

1. Dang, T. & Süssmuth, R. D. Bioactive Peptide Natural Products as Lead Structures for Medicinal Use. *Acc Chem Res* **50,** 1566–1576 (2017).
2. Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discov Today* **15,** 40–56 (2010).
3. Danho, W., Swistok, J., Khan, W., Chu, X.-J., Cheung, A., Fry, D., Sun, H., Kurylko, G., Rumennik, L., Cefalu, J., *et al.* Opportunities and challenges of developing peptide drugs in the pharmaceutical industry. *Adv Exp Med Biol* **611,** 467–9 (2009).
4. Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Schäberle, T. F., Hughes, D. E., Epstein, S., *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517,** 455–9 (2015).
5. Hover, B. M., Kim, S.-H., Katz, M., Charlop-Powers, Z., Owen, J. G., Ternei, M. A., Maniko, J., Estrela, A. B., Molina, H., Park, S., *et al.* Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat Microbiol* **3,** 415–422 (2018).
6. Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides1. *Annu Rev Microbiol* **58,** 453–88 (2004).
7. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew Chem Int Ed Engl* **56,** 3770–3821 (2017).
8. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat Prod Rep* **33,** 317–47 (2016).
9. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc Natl Acad Sci U S A* **111,** 9259–64 (2014).
10. Richardt, A., Kemme, T., Wagner, S., Schwarzer, D., Marahiel, M. A. & Hovemann, B. T. Ebony, a novel nonribosomal peptide synthetase for beta-alanine conjugation with biogenic amines in Drosophila. *J Biol Chem* **278,** 41160–6 (2003).

11. Shou, Q., Feng, L., Long, Y., Han, J., Nunnery, J. K., Powell, D. H. & Butcher, R. A. A hybrid polyketide-nonribosomal peptide in nematodes that promotes larval survival. *Nat Chem Biol* **12,** 770–2 (Oct. 2016).

12. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol* **9,** 522–6 (2002).

13. Healy, F. G., Wach, M, Krasnoff, S. B., Gibson, D. M. & Loria, R. The txtAB genes of the plant pathogen Streptomyces acidiscabies encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol Microbiol* **38,** 794–804 (2000).

14. Molnár, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D. D., *et al.* The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. *Chem Biol* **7,** 97–109 (2000).

15. Schmidt, B., Riesner, D., Lawen, A. & Kleinkauf, H. Cyclosporin synthetase is a 1.4 MDa multienzyme polypeptide. Re-evaluation of the molecular mass of various peptide synthetases. *FEBS Lett* **307,** 355–60 (1992).

16. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30,** 108–60 (2013).

17. Mohr, K. I., Volz, C., Jansen, R., Wray, V., Hoffmann, J., Bernecker, S., Wink, J., Gerth, K., Stadler, M. & Müller, R. Pinensins: the first antifungal lantibiotics. *Angew Chem Int Ed Engl* **54,** 11254–8 (2015).

18. Férir, G., Petrova, M. I., Andrei, G., Huskens, D., Hoorelbeke, B., Snoeck, R., Vanderleyden, J., Balzarini, J., Bartoschek, S., Brönstrup, M., *et al.* The lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *PLoS One* **8,** e64010 (2013).

19. Knerr, P. J. & van der Donk, W. A. Discovery, biosynthesis, and engineering of lantipeptides. *Annu Rev Biochem* **81,** 479–505 (2012).

20. Martins, J. & Vasconcelos, V. Cyanobactins from Cyanobacteria: Current Genetic and Chemical State of Knowledge. *Mar Drugs* **13,** 6910–46 (2015).

21. Sivonen, K., Leikoski, N., Fewer, D. P. & Jokela, J. Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Appl Microbiol Biotechnol* **86,** 1213–25 (2010).

22. Just-Baringo, X., Albericio, F. & Álvarez, M. Thiopeptide antibiotics: retrospective and recent advances. *Mar Drugs* **12,** 317–51 (2014).

23. McIntosh, J. A., Donia, M. S. & Schmidt, E. W. Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. *Nat Prod Rep* **26,** 537–59 (2009).

24. Wilkinson, B. & Micklefield, J. Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* **3,** 379–86 (2007).

25. Williams, G. J. Engineering polyketide synthases and nonribosomal peptide synthetases. *Curr Opin Struct Biol* **23,** 603–12 (2013).

26. Hetrick, K. J. & van der Donk, W. A. Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Curr Opin Chem Biol* **38,** 36–44 (2017).

27. Schnell, N., Entian, K. D., Schneider, U., Götz, F., Zähner, H., Kellner, R. & Jung, G. Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. *Nature* **333,** 276–8 (1988).

28. Ortega, M. A. & van der Donk, W. A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem Biol* **23,** 31–44 (2016).

29. Thibodeaux, G. N., McClerren, A. L., Ma, Y., Gancayco, M. R. & van der Donk, W. A. Synergistic binding of the leader and core peptides by the lantibiotic synthetase HalM2. *ACS Chem Biol* **10,** 970–7 (2015).

30. Burkhart, B. J., Kakkar, N., Hudson, G. A., van der Donk, W. A. & Mitchell, D. A. Chimeric Leader Peptides for the Generation of Non-Natural Hybrid RiPP Products. *ACS Cent Sci* **3,** 629–638 (2017).

31. Williams, D. E., Moore, R. E. & Paul, V. J. The structure of ulithiacyclamide B. Antitumor evaluation of cyclic peptides and macrolides from Lissoclinum patella. *J Nat Prod* **52,** 732–9 (1989).

32. Rashid, M. A., Gustafson, K. R., Cardellina, J. H. 2nd & Boyd, M. R. Patellamide F, A new cytotoxic cyclic peptide from the colonial ascidian Lissoclinum patella. *J Nat Prod* **58,** 594–7 (1995).

33. Fu, X, Do, T, Schmitz, F. J., Andrusevich, V & Engel, M. H. New cyclic peptides from the ascidian Lissoclinum patella. *J Nat Prod* **61,** 1547–51 (1998).

34. Linington, R. G., Gonzalez, J., Urena, L.-D., Romero, L. I., Ortega-Barria, E. & Gerwick, W. H. Venturamides A and B: Antimalarial constituents of the Panamanian marine cyanobacterium Oscillatoria sp. *J. Nat. Prod.* **70,** 397–401 (2007).

35. Portmann, C., Blom, J. F., Kaiser, M., Brun, R., Jüttner, F. & Gademann, K. Isolation of aerucyclamides C and D and structure revision of microcyclamide 7806A: heterocyclic ribosomal peptides from Microcystis aeruginosa PCC 7806 and their antiparasite evaluation. *J Nat Prod* **71,** 1891–6 (2008).

36. Ireland, C. & Scheuer, P. J. Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate. *Journal of the American Chemical Society* **102,** 5688–5691 (1980).

37. Long, P. F., Dunlap, W. C., Battershill, C. N. & Jaspars, M. Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. *Chembiochem* **6,** 1760–5 (2005).

38. Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G. & Ravel, J. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 7315–7320 (2005).

39. Schmidt, E. W. & Donia, M. S. Chapter 23. Cyanobactin ribosomally synthesized peptides–a case of deep metagenome mining. *Methods Enzymol* **458,** 575–96 (2009).

40. Leikoski, N., Liu, L., Jokela, J., Wahlsten, M., Gugger, M., Calteau, A., Permi, P., Kerfeld, C. A., Sivonen, K. & Fewer, D. P. Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. *Chem Biol* **20,** 1033–43 (2013).

41. Wahyudi, H. & McAlpine, S. R. Predicting the unpredictable: Recent structure-activity studies on peptide-based macrocycles. *Bioorg Chem* **60,** 74–97 (2015).

42. García-Reynaga, P. & VanNieuwenhze, M. S. A new total synthesis of patellamide A. *Org Lett* **10,** 4621–3 (2008).

43. You, S.-L., Razavi, H. & Kelly, J. W. A biomimetic synthesis of thiazolines using hexaphenyloxodiphosphonium trifluoromethanesulfonate. *Angew Chem Int Ed Engl* **42,** 83–5 (2003).

44. Koehnke, J., Bent, A. F., Houssen, W. E., Mann, G., Jaspars, M. & Naismith, J. H. The structural biology of patellamide biosynthesis. *Curr Opin Struct Biol* **29,** 112–21 (2014).

45. Czekster, C. M., Ge, Y. & Naismith, J. H. Mechanisms of cyanobactin biosynthesis. *Curr Opin Chem Biol* **35,** 80–88 (2016).

46. Salvatella, X., Caba, J. M., Albericio, F. & Giralt, E. Solution structure of the antitumor candidate trunkamide A by 2D NMR and restrained simulated annealing methods. *J Org Chem* **68,** 211–5 (2003).

47. Wipf, P & Uto, Y. Total synthesis and revision of stereochemistry of the marine metabolite trunkamide A. *J Org Chem* **65,** 1037–49 (2000).

48. Ishida, K, Nakagawa, H & Murakami, M. Microcyclamide, a cytotoxic cyclic hexapeptide from the cyanobacterium Microcystis aeruginosa. *J Nat Prod* **63,** 1315–7 (2000).

49. Ziemert, N., Ishida, K., Quillardet, P., Bouchier, C., Hertweck, C., de Marsac, N. T. & Dittmann, E. Microcyclamide biosynthesis in two strains of Microcystis aeruginosa: from structure to genes and vice versa. *Appl Environ Microbiol* **74,** 1791–7 (2008).

50. Donia, M. S., Ravel, J. & Schmidt, E. W. A global assembly line for cyanobactins. *Nat Chem Biol* **4,** 341–3 (2008).

51. Martins, J., Leão, P. N., Ramos, V. & Vasconcelos, V. N-terminal protease gene phylogeny reveals the potential for novel cyanobactin diversity in cyanobacteria. *Mar Drugs* **11,** 4902–16 (2013).

52. Roy, R. S., Gehring, A. M., Milne, J. C., Belshaw, P. J. & Walsh, C. T. Thiazole and oxazole peptides: biosynthesis and molecular machinery. *Nat Prod Rep* **16,** 249–63 (1999).

53. Hartmut, D., Holger, S., Ralf, L., Hubert, H., Hans, Z., Klaus, H. & Günther, J. Structure elucidation of yersiniabactin, a siderophore from highly virulent Yersinia strains. *Liebigs Annalen* **1995,** 1727–1733.

54. Snow, G. A. Mycobactins: iron-chelating growth factors from mycobacteria. *Bacteriol Rev* **34,** 99–125 (1970).

55. Barclay, R & Ratledge, C. Mycobactins and exochelins of Mycobacterium tuberculosis, M. bovis, M. africanum and other related species. *J Gen Microbiol* **134,** 771–6 (1988).

56. Griffiths, G. L., Sigel, S. P., Payne, S. M. & Neilands, J. B. Vibriobactin, a siderophore from Vibrio cholerae. *J Biol Chem* **259,** 383–5 (1984).

57. Lindquist, N., Fenical, W., Van Duyne, G. D. & Clardy, J. Isolation and structure determination of diazonamides A and B, unusual cytotoxic metabolites from the marine ascidian Diazona chinensis. *Journal of the American Chemical Society* **113,** 2303–2304 (1991).

58. Kunze, B, Jansen, R, Pridzun, L, Jurkiewicz, E, Hunsmann, G, Höfle, G & Reichenbach, H. Thiangazole, a new thiazoline antibiotic from Polyangium sp. (myxobacteria): production, antimicrobial activity and mechanism of action. *J Antibiot (Tokyo)* **46,** 1752–5 (1993).

59. Jin, Z. Muscarine, imidazole, oxazole and thiazole alkaloids. *Nat Prod Rep* **33,** 1268–1317 (2016).

60. Storm, D. R. & Strominger, J. L. Complex formation between bacitracin peptides and isoprenyl pyrophosphates. The specificity of lipid-peptide interactions. *J Biol Chem* **248,** 3940–5 (1973).

61. Vanderwall, D. E., Lui, S. M., Wu, W, Turner, C. J., Kozarich, J. W. & Stubbe, J. A model of the structure of HOO-Co.bleomycin bound to d(CCAGTACTGG): recognition at the d(GpT) site and implications for double-stranded DNA cleavage. *Chem Biol* **4,** 373–87 (1997).

62. Selva, E, Montanini, N, Stella, S, Soffientini, A, Gastaldo, L & Denaro, M. Targeted screening for elongation factor Tu binding antibiotics. *J Antibiot (Tokyo)* **50,** 22–6 (1997).

63. Anderson, B, Hodgkin, D. C. & Viswamitra, M. A. The structure of thiostrepton. *Nature* **225,** 233–5 (1970).

64. Koehnke, J., Mann, G., Bent, A. F., Ludewig, H., Shirran, S., Botting, C., Lebl, T., Houssen, W. E., Jaspars, M. & Naismith, J. H. Structural analysis of leader peptide binding enables leader-free cyanobactin processing. *Nat Chem Biol* (2015).

65. Koehnke, J., Bent, A. F., Zollman, D., Smith, K., Houssen, W. E., Zhu, X., Mann, G., Lebl, T., Scharff, R., Shirran, S., *et al.* The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. *Angew Chem Int Ed Engl* **52,** 13991–6 (2013).

66. McIntosh, J. A. & Schmidt, E. W. Marine molecular machines: heterocyclization in cyanobactin biosynthesis. *Chembiochem* **11,** 1413–21 (2010).

67. McIntosh, J. A., Donia, M. S. & Schmidt, E. W. Insights into heterocyclization from two highly similar enzymes. *J Am Chem Soc* **132,** 4089–91 (2010).

68. Melby, J. O., Dunbar, K. L., Trinh, N. Q. & Mitchell, D. A. Selectivity, directionality, and promiscuity in peptide processing from a Bacillus sp. Al Hakam cyclodehydratase. *J Am Chem Soc* **134,** 5309–16 (2012).

69. Burkhart, B. J., Schwalen, C. J., Mann, G., Naismith, J. H. & Mitchell, D. A. YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem Rev* **117,** 5389–5456 (2017).

70. Li, Y. M., Milne, J. C., Madison, L. L., Kolter, R & Walsh, C. T. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science* **274,** 1188–93 (1996).

71. Dunbar, K. L., Melby, J. O. & Mitchell, D. A. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat Chem Biol* **8,** 569–75 (2012).

72. Dunbar, K. L., Chekan, J. R., Cox, C. L., Burkhart, B. J., Nair, S. K. & Mitchell, D. A. Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat Chem Biol* **10,** 823–9 (2014).

73. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim Biophys Acta* **1865,** 1587–1604 (Nov. 2017).

74. Bloudoff, K., Fage, C. D., Marahiel, M. A. & Schmeing, T. M. Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc Natl Acad Sci U S A* **114,** 95–100 (Jan. 2017).

75. Dowling, D. P., Kung, Y., Croft, A. K., Taghizadeh, K., Kelly, W. L., Walsh, C. T. & Drennan, C. L. Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc Natl Acad Sci U S A* **113,** 12432–12437 (Nov. 2016).

76. Konz, D., Klens, A., Schörgendorfer, K. & Marahiel, M. A. The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* **4,** 927–37 (1997).

77. Goto, Y., Ito, Y., Kato, Y., Tsunoda, S. & Suga, H. One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase. *Chem Biol* **21,** 766–74 (2014).

78. Oman, T. J. & van der Donk, W. A. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol* **6,** 9–18 (2010).

79. Burkhart, B. J., Hudson, G. A., Dunbar, K. L. & Mitchell, D. A. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat Chem Biol* (2015).

80. Ortega, M. A., Hao, Y., Zhang, Q., Walker, M. C., van der Donk, W. A. & Nair, S. K. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* **517,** 509–12 (2015).

81. Khusainov, R. & Kuipers, O. P. When the leader gets loose: in vivo biosynthesis of a leaderless prenisin is stimulated by a trans-acting leader peptide. *Chembiochem* **13,** 2433–8 (2012).

82. Patton, G. C., Paul, M., Cooper, L. E., Chatterjee, C. & van der Donk, W. A. The importance of the leader sequence for directing lanthionine formation in lacticin 481. *Biochemistry* **47,** 7342–51 (2008).

83. Oman, T. J., Knerr, P. J., Bindman, N. A., Velásquez, J. E. & van der Donk, W. A. An engineered lantibiotic synthetase that does not require a leader peptide on its substrate. *J Am Chem Soc* **134,** 6952–5 (2012).

84. Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* **1079,** 105–16 (2014).

85. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42,** W320–4 (2014).

86. Rudi, A., Aknin, M., Gaydou, E. M. & Kashman, Y. Four new cytotoxic cyclic hexa- and heptapeptides from the marine ascidian Didemnum molle. *Tetrahedron* **54,** 13203 –13210 (1998).

87. Houssen, W. E., Bent, A. F., McEwan, A. R., Pieiller, N., Tabudravu, J., Koehnke, J., Mann, G., Adaba, R. I., Thomas, L., Hawas, U. W., *et al.* An efficient method for the in vitro production of azol(in)e-based cyclic peptides. *Angew Chem Int Ed Engl* **53,** 14171–4 (2014).

88. Thibodeaux, C. J., Wagoner, J., Yu, Y. & van der Donk, W. A. Leader Peptide Establishes Dehydration Order, Promotes Efficiency, and Ensures Fidelity During Lacticin 481 Biosynthesis. *J Am Chem Soc* **138,** 6436–44 (May 2016).

89. Lee, M. V., Ihnken, L. A. F., You, Y. O., McClerren, A. L., van der Donk, W. A. & Kelleher, N. L. Distributive and directional behavior of lantibiotic synthetases revealed by high-resolution tandem mass spectrometry. *J Am Chem Soc* **131,** 12258–64 (2009).

90. Agarwal, V., Pierce, E., McIntosh, J., Schmidt, E. W. & Nair, S. K. Structures of cyanobactin maturation enzymes define a family of transamidating proteases. *Chem Biol* **19,** 1411–22 (2012).

91. Koehnke, J., Bent, A., Houssen, W. E., Zollman, D., Morawitz, F., Shirran, S., Vendome, J., Nneoyiegbe, A. F., Trembleau, L., Botting, C. H., *et al.* The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nat Struct Mol Biol* **19,** 767–72 (2012).

92. Sardar, D., Lin, Z. & Schmidt, E. W. Modularity of RiPP Enzymes Enables Designed Synthesis of Decorated Peptides. *Chemistry & Biology* (2015).

93. Ruffner, D. E., Schmidt, E. W. & Heemstra, J. R. Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. *ACS Synth Biol* **4,** 482–92 (2015).

94. Oueis, E., Adamson, C., Mann, G., Ludewig, H., Redpath, P., Migaud, M., Westwood, N. J. & Naismith, J. H. Derivatisable Cyanobactin Analogues: A Semisynthetic Approach. *Chembiochem* **16,** 2646–50 (2015).

95. Oueis, E., Nardone, B., Jaspars, M., Westwood, N. J. & Naismith, J. H. Synthesis of Hybrid Cyclopeptides through Enzymatic Macrocyclization. *ChemistryOpen* **6,** 11–14 (Feb. 2017).

96. Oueis, E., Jaspars, M., Westwood, N. J. & Naismith, J. H. Enzymatic Macrocyclization of 1,2,3-Triazole Peptide Mimetics. *Angew Chem Int Ed Engl* **55,** 5842–5 (2016).

97. Thévenet, P., Shen, Y., Maupetit, J., Guyon, F., Derreumaux, P. & Tufféry, P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res* **40,** W288–93 (2012).

98. Shen, Y., Maupetit, J., Derreumaux, P. & Tufféry, P. Improved PEP-FOLD Approach for Peptide and Miniprotein Structure Prediction. *J Chem Theory Comput* **10,** 4745–58 (2014).

99. Lee, J., McIntosh, J., Hathaway, B. J. & Schmidt, E. W. Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates. *J Am Chem Soc* **131,** 2122–4 (2009).

100. McIntosh, J. A., Robertson, C. R., Agarwal, V., Nair, S. K., Bulaj, G. W. & Schmidt, E. W. Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. *J Am Chem Soc* **132,** 15499–501 (2010).

101. Tianero, M. D., Pierce, E., Raghuraman, S., Sardar, D., McIntosh, J. A., Heemstra, J. R., Schonrock, Z., Covington, B. C., Maschek, J. A., Cox, J. E., *et al.* Metabolic model for diversity-generating biosynthesis. *Proc Natl Acad Sci U S A* **113,** 1772–7 (2016).

102. Luo, H., Hong, S.-Y., Sgambelluri, R. M., Angelos, E., Li, X. & Walton, J. D. Peptide macrocyclization catalyzed by a prolyl oligopeptidase involved in -amanitin biosynthesis. *Chem Biol* **21,** 1610–7 (2014).

103. Barber, C. J. S., Pujara, P. T., Reed, D. W., Chiwocha, S., Zhang, H. & Covello, P. S. The Two-step Biosynthesis of Cyclic Peptides from Linear Precursors in a Member of the Plant Family Caryophyllaceae Involves Cyclization by a Serine Protease-like Enzyme. *Journal of Biological Chemistry* **288,** 12500–12510 (2013).

104. Nguyen, G. K. T., Wang, S., Qiu, Y., Hemu, X., Lian, Y. & Tam, J. P. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat Chem Biol* **10,** 732–8 (2014).

105. Nguyen, G. K. T., Kam, A., Loo, S., Jansson, A. E., Pan, L. X. & Tam, J. P. Butelase 1: A Versatile Ligase for Peptide and Protein Macrocyclization. *J Am Chem Soc* **137,** 15398–401 (2015).

106. Bent, A. F., Mann, G., Houssen, W. E., Mykhaylyk, V., Duman, R., Thomas, L., Jaspars, M., Wagner, A. & Naismith, J. H. Structure of the cyanobactin oxidase ThcOx from Cyanothece sp. PCC 7425, the first structure to be solved at Diamond Light Source beamline I23 by means of S-SAD. *Acta Crystallogr D Struct Biol* **72,** 1174–1180 (Nov. 2016).

107. Katritzky, A. R., Ramsden, C. A., A., J. J. & V., Z. V. *Handbook of Heterocyclic Chemistry* 3rd ed. (Oxford : Elsevier, 2010).

108. Blackburn, S., Middlebrook, W. R. & Phillips, H. Oxazoline and Thiazoline Rings in Proteins. *Nature* **150,** 57 (July 1942).

109. Marques, S. M. & Esteves da Silva, J. C. G. Firefly bioluminescence: a mechanistic approach of luciferase catalyzed reactions. *IUBMB Life* **61,** 6–17 (2009).

110. Baryck, J. J., Asard, H., Stone, J. M., Wilson, M. A., Banerjee, R. & Becker, D. F. in *Redox Biochemistry* 11–47 (Wiley-Blackwell, 2007).

111. Van Vranken, J. G., Na, U., Winge, D. R. & Rutter, J. Protein-mediated assembly of succinate dehydrogenase and its cofactors. *Crit Rev Biochem Mol Biol* **50,** 168–80 (2015).

112. Dunbar, K. L. & Mitchell, D. A. Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. *J Am Chem Soc* **135,** 8692–701 (2013).

113. Melby, J. O., Li, X. & Mitchell, D. A. Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry* **53,** 413–22 (2014).

114. Schneider, T. L., Shen, B. & Walsh, C. T. Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42,** 9722–30 (2003).

115. Milne, J. C., Roy, R. S., Eliot, A. C., Kelleher, N. L., Wokhlu, A, Nickels, B & Walsh, C. T. Cofactor requirements and reconstitution of microcin B17 synthetase: a multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17. *Biochemistry* **38,** 4768–81 (1999).

116. Gonzalez, D. J., Lee, S. W., Hensler, M. E., Markley, A. L., Dahesh, S., Mitchell, D. A., Bandeira, N., Nizet, V., Dixon, J. E. & Dorrestein, P. C. Clostridiolysin S, a post-translationally modified biotoxin from Clostridium botulinum. *J Biol Chem* **285,** 28220–8 (2010).

117. Zverina, E. A., Lamphear, C. L., Wright, E. N. & Fierke, C. A. Recent advances in protein prenyltransferases: substrate identification, regulation, and disease interventions. *Curr Opin Chem Biol* **16,** 544–52 (2012).

118. McIntosh, J. A., Lin, Z., Tianero, M. D. B. & Schmidt, E. W. Aestuaramides, a natural library of cyanobactin cyclic peptides resulting from isoprene-derived Claisen rearrangements. *ACS Chem Biol* **8,** 877–83 (2013).

119. Donia, M. S. & Schmidt, E. W. Linking chemistry and genetics in the growing cyanobactin natural products family. *Chem Biol* **18,** 508–19 (2011).

120. Donia, M. S., Fricke, W. F., Ravel, J. & Schmidt, E. W. Variation in tropical reef symbiont metagenomes defined by secondary metabolism. *PLoS One* **6,** e17897 (2011).

121. Majmudar, J. D. & Gibbs, R. A. Pericyclic prenylation: peptide modification through a Claisen rearrangement. *Chembiochem* **12,** 2723–6 (2011).

122. McIntosh, J. A., Donia, M. S., Nair, S. K. & Schmidt, E. W. Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. *J Am Chem Soc* **133,** 13698–705 (2011).

123. Osuna, S., Kim, S., Bollot, G. & Houk, K. N. Aromatic Claisen Rearrangements of O-prenylated tyrosine and model prenyl aryl ethers: Computational study of the role of water on acceleration of Claisen rearrangements. *European J Org Chem* **2013** (2013).

124. Parajuli, A., Kwak, D. H., Dalponte, L., Leikoski, N., Galica, T., Umeobika, U., Trembleau, L., Bent, A., Sivonen, K., Wahlsten, M., *et al.* A Unique Tryptophan C-Prenyltransferase from the Kawaguchipeptin Biosynthetic Pathway. *Angew Chem Int Ed Engl* **55,** 3596–9 (2016).

125. Leikoski, N., Fewer, D. P., Jokela, J., Alakoski, P., Wahlsten, M. & Sivonen, K. Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus Microcystis. *PLoS One* **7,** e43002 (2012).

126. Sardar, D., Hao, Y., Lin, Z., Morita, M., Nair, S. K. & Schmidt, E. W. Enzymatic N- and C-Protection in Cyanobactin RiPP Natural Products. *J Am Chem Soc* **139,** 2884–2887 (Mar. 2017).

127. Hao, Y., Pierce, E., Roe, D., Morita, M., McIntosh, J. A., Agarwal, V., Cheatham, T. E. 3rd, Schmidt, E. W. & Nair, S. K. Molecular basis for the broad substrate selectivity of a peptide prenyltransferase. *Proc Natl Acad Sci U S A* **113,** 14037–14042 (Dec. 2016).

128. Morita, M., Hao, Y., Jokela, J. K., Sardar, D., Lin, Z., Sivonen, K., Nair, S. K. & Schmidt, E. W. Post-Translational Tyrosine Geranylation in Cyanobactin Biosynthesis. *J Am Chem Soc* **140,** 6044–6048 (May 2018).

129. Leikoski, N., Fewer, D. P., Jokela, J., Wahlsten, M., Rouhiainen, L. & Sivonen, K. Highly diverse cyanobactins in strains of the genus Anabaena. *Appl Environ Microbiol* **76,** 701–9 (2010).

130. Houssen, W. E., Koehnke, J., Zollman, D., Vendome, J., Raab, A., Smith, M. C. M., Naismith, J. H. & Jaspars, M. The discovery of new cyanobactins from Cyanothece PCC 7425 defines a new signature for processing of patellamides. *Chembiochem* **13,** 2683–9 (2012).

131. Banker, R & Carmeli, S. Tenuecyclamides A-D, cyclic hexapeptides from the cyanobacterium Nostoc spongiaeforme var. tenue. *J Nat Prod* **61,** 1248–51 (1998).

132. Bent, A. F., Koehnke, J., Houssen, W. E., Smith, M. C. M., Jaspars, M. & Naismith, J. H. Structure of PatF from Prochloron didemni. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **69,** 618–23 (2013).

133. Metzger, U., Keller, S., Stevenson, C. E. M., Heide, L. & Lawson, D. M. Structure and mechanism of the magnesium-independent aromatic prenyltransferase CloQ from the clorobiocin biosynthetic pathway. *J Mol Biol* **404,** 611–26 (2010).

134. Metzger, U., Schall, C., Zocher, G., Unsöld, I., Stec, E., Li, S.-M., Heide, L. & Stehle, T. The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. *Proc Natl Acad Sci U S A* **106,** 14309–14 (2009).

135. Kuzuyama, T., Noel, J. P. & Richard, S. B. Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature* **435,** 983–7 (2005).

136. Milne, B. F., Long, P. F., Starcevic, A., Hranueli, D. & Jaspars, M. Spontaneity in the patellamide biosynthetic pathway. *Org Biomol Chem* **4,** 631–8 (2006).

137. Tianero, M. D. B., Kwan, J. C., Wyche, T. P., Presson, A. P., Koch, M., Barrows, L. R., Bugni, T. S. & Schmidt, E. W. Species specificity of symbiosis and secondary metabolism in ascidians. *ISME J* **9,** 615–28 (2015).

138. Schmidt, E. W. The secret to a successful relationship: lasting chemistry between ascidians and their symbiotic bacteria. *Invertebr Biol* **134,** 88–102 (2015).

139. Tianero, M. D. B., Donia, M. S., Young, T. S., Schultz, P. G. & Schmidt, E. W. Ribosomal route to small-molecule diversity. *J Am Chem Soc* **134,** 418–25 (2012).

140. Sardar, D., Pierce, E., McIntosh, J. A. & Schmidt, E. W. Recognition sequences and substrate evolution in cyanobactin biosynthesis. *ACS Synth Biol* **4,** 167–76 (2015).

141. Plat, A., Kluskens, L. D., Kuipers, A., Rink, R. & Moll, G. N. Requirements of the engineered leader peptide of nisin for inducing modification, export, and cleavage. *Appl Environ Microbiol* **77,** 604–11 (2011).

142. Bindman, N. A., Bobeica, S. C., Liu, W. R. & van der Donk, W. A. Facile Removal of Leader Peptides from Lanthipeptides by Incorporation of a Hydroxy Acid. *J Am Chem Soc* **137,** 6975–8 (2015).

143. Koehnke, J., Morawitz, F., Bent, A. F., Houssen, W. E., Shirran, S. L., Fuszard, M. A., Smellie, I. A., Botting, C. H., Smith, M. C. M., Jaspars, M., *et al.* An enzymatic route to selenazolines. *Chembiochem* **14,** 564–7 (2013).

144. Sambrook, J. & Russell, D. W. The inoue method for preparation and transformation of competent e. Coli: "ultra-competent" cells. *CSH Protoc* **2006** (2006).

145. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8,** 91 (2008).

146. Liu, H. & Naismith, J. H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **63,** 102–11 (2009).

147. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41,** 207–34 (2005).

148. Wilkins, M. R., Gasteiger, E, Bairoch, A, Sanchez, J. C., Williams, K. L., Appel, R. D. & Hochstrasser, D. F. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112,** 531–52 (1999).

149. Polson, C., Sarkar, P., Incledon, B., Raguvaran, V. & Grant, R. Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **785,** 263–75 (2003).

150. Morrison, J. F. Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim Biophys Acta* **185,** 269–86 (1969).

151. Wang, P. & Giese, R. W. Recommendations for quantitative analysis of small molecules by matrix-assisted laser desorption ionization mass spectrometry. *J Chromatogr A* **1486,** 35–41 (2017).

152. Repka, L. M., Hetrick, K. J., Chee, S. H. & van der Donk, W. A. Characterization of Leader Peptide Binding During Catalysis by the Nisin Dehydratase NisB. *J Am Chem Soc* **140,** 4200–4203 (2018).

153. Lubelski, J., Khusainov, R. & Kuipers, O. P. Directionality and coordination of dehydration and ring formation during biosynthesis of the lantibiotic nisin. *J Biol Chem* **284,** 25962–72 (2009).

154. Krawczyk, B., Ensle, P., Müller, W. M. & Süssmuth, R. D. Deuterium labeled peptides give insights into the directionality of class III lantibiotic synthetase LabKC. *J Am Chem Soc* **134,** 9922–5 (2012).

155. Kelleher, N. L., Belshaw, P. J. & Walsh, C. T. Regioselectivity and Chemoselectivity Analysis of Oxazole and Thiazole Ring Formation by the Peptide-Heterocyclizing Microcin B17 Synthetase Using High-Resolution MS/MS. *Journal of the American Chemical Society* **120,** 9716–9717 (1998).

156. Mukherjee, S. & van der Donk, W. A. Mechanistic studies on the substrate-tolerant lanthipeptide synthetase ProcM. *J Am Chem Soc* **136,** 10450–9 (2014).

157. Miller, J. H. *Experiments in molecular genetics* (Cold Spring Harbor Laboratory, 1972).

158. Delaglio, F, Grzesiek, S, Vuister, G. W., Zhu, G, Pfeifer, J & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6,** 277–93 (1995).

159. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–96 (2005).

160. Grzesiek, S. & Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **114,** 6291–6293 (1992).

161. Prabhu, V., Chatson, B., Abrams, G. & King, J. 13C Chemical shifts of 20 free amino acids and their use in detection by NMR of free amino acids in intact plants of Arabidopsis. *Journal of Plant Physiology* **149,** 246 –250 (1996).

162. Thibodeaux, C. J., Ha, T. & van der Donk, W. A. A price to pay for relaxed substrate specificity: a comparative kinetic analysis of the class II lanthipeptide synthetases ProcM and HalM2. *J Am Chem Soc* **136,** 17513–29 (2014).

163. Ludewig, H., Czekster, C. M., Oueis, E., Munday, E. S., Arshad, M., Synowsky, S. A., Bent, A. F. & Naismith, J. H. Characterization of the Fast and Promiscuous Macrocyclase from Plant PCY1 Enables the Use of Simple Substrates. *ACS Chem Biol* **13,** 801–811 (2018).

164. Vauquelin, G. & Charlton, S. J. Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *Br J Pharmacol* **168,** 1771–85 (2013).

165. Di Cera, E. Chapter 11 - Kinetics of Allosteric Activation. *Methods in Enzymology* **466,** 259–271 (2009).

166. Hawkins, C. J., Lavin, M. F., Marshall, K. A., van den Brenk, A. L. & Watters, D. J. Structure-activity relationships of the lissoclinamides: cytotoxic cyclic peptides from the ascidian Lissoclinum patella. *J Med Chem* **33,** 1634–8 (1990).

167. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33,** W244–8 (2005).

168. Gao, S., Ge, Y., Bent, A. F., Schwarz-Linek, U. & Naismith, J. H. Oxidation of cyanobactin precursor peptide is independent of leader peptide and operates in a defined order. *Biochemistry.* PMID: 30208270.

169. Davis, K. M., Schramma, K. R., Hansen, W. A., Bacik, J. P., Khare, S. D., Seyedsayamdost, M. R. & Ando, N. Structures of the peptide-modifying radical SAM enzyme SuiB elucidate the basis of substrate recognition. *Proc Natl Acad Sci U S A* **114,** 10420–10425 (2017).

170. Thompson, P. R. & Cole, P. A. Probing the mechanism of enzymatic phosphoryl transfer with a chemical trick. *Proc Natl Acad Sci U S A* **98,** 8170–1 (2001).

171. Lassila, J. K., Zalatan, J. G. & Herschlag, D. Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annu Rev Biochem* **80,** 669–702 (2011).

172. FINK, G. R. GENE-ENZYME RELATIONS IN HISTIDINE BIOSYNTHESIS IN YEAST. *Science* **146,** 525–7 (1964).

173. Chapman, L. F. & Nester, E. W. Gene-enzyme relationships in histidine biosynthesis in Bacillus subtilis. *J Bacteriol* **97,** 1444–8 (1969).

174. Stepansky, A & Leustek, T. Histidine biosynthesis in plants. *Amino Acids* **30,** 127–42 (2006).

175. Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V. & Dixon, J. E. Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc Natl Acad Sci U S A* **105,** 5879–84 (2008).

176. AMES, B. N., MARTIN, R. G. & GARRY, B. J. The first step of histidine biosynthesis. *J Biol Chem* **236,** 2019–26 (1961).

177. Alifano, P, Fani, R, Liò, P, Lazcano, A, Bazzicalupo, M, Carlomagno, M. S. & Bruni, C. B. Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol Rev* **60,** 44–69 (1996).

178. Kulis-Horn, R. K., Persicke, M. & Kalinowski, J. Histidine biosynthesis, its regulation and biotechnological application in Corynebacterium glutamicum. *Microb Biotechnol* **7,** 5–25 (2014).

179. Martin, R. G. The First Enzyme in Histidine Biosynthesis: The Nature of Feedback Inhibition by Histidine. *Journal of Biological Chemistry* **238,** 257–268 (Jan. 1963).

180. Morton, D. P. & Parsons, S. M. Inhibition of ATP phosphoribosyltransferase by AMP and ADP in the absence and presence of histidine. *Arch Biochem Biophys* **181,** 643–8 (1977).

181. Rees, J. D., Ingle, R. A. & Smith, J. A. C. Relative contributions of nine genes in the pathway of histidine biosynthesis to the control of free histidine concentrations in Arabidopsis thaliana. *Plant Biotechnol J* **7,** 499–511 (2009).

182. Blasi, F, Aloj, S. M. & Goldberger, R. F. Effect of histidine on the enzyme which catalyzes the first step of histidine biosynthesis in Salmonella typhimurium. *Biochemistry* **10,** 1409–17 (1971).

183. Blasi, F, Barton, R. W., Kovach, J. S. & Goldberger, R. F. Interaction between the first enzyme for histidine biosynthesis and histidyl transfer ribonucleic acid. *J Bacteriol* **106,** 508–13 (1971).

184. Brashear, W. T. & Parsons, S. M. Evidence against a covalent intermediate in the adenosine triphosphate phosphoribosyltransferase reaction of histidine biosynthesis. *J Biol Chem* **250,** 6885–90 (1975).

185. Kleeman, J. E. & Parsons, S. M. Reverse direction substrate kinetics and inhibition studies on the first enzyme of histidine biosynthesis, adenosine triphosphate phosphoribosyltransferase. *Arch Biochem Biophys* **175,** 687–93 (1976).

186. Morton, D. P. & Parsons, S. M. Biosynthetic direction substrate kinetics and product inhibition studies on the first enzyme of histidine biosynthesis, adenosine triphosphate phosphoribosyltransferase. *Arch Biochem Biophys* **175,** 677–86 (1976).

187. Dall-Larsen, T & Klungsoyr, L. The binding of specific ligands to adenosine-triphosphate phosphoribosyltransferase. *Eur J Biochem* **69,** 195–201 (1976).

188. Pedreño, S., Pisco, J. P., Larrouy-Maumus, G., Kelly, G. & de Carvalho, L. P. S. Mechanism of feedback allosteric inhibition of ATP phosphoribosyltransferase. *Biochemistry* **51,** 8027–38 (2012).

189. Mittelstädt, G., Moggré, G.-J., Panjikar, S., Nazmi, A. R. & Parker, E. J. Campylobacter jejuni adenosine triphosphate phosphoribosyltransferase is an active hexamer that is allosterically controlled by the twisting of a regulatory tail. *Protein Sci* **25,** 1492–506 (Aug. 2016).

190. Pacholarz, K. J., Burnley, R. J., Jowitt, T. A., Ordsmith, V., Pisco, J. P., Porrini, M., Larrouy-Maumus, G., Garlish, R. A., Taylor, R. J., de Carvalho, L. P. S., *et al.* Hybrid Mass Spectrometry Approaches to Determine How L-Histidine Feedback Regulates the Enzyzme MtATP-Phosphoribosyltransferase. *Structure* **25,** 730–738.e4 (2017).

191. Pisco, J. P., Chiara, C. d., Pacholarz, K. J., Garza-Garcia, A., Ogrodowicz, R. W., Walker, P. A., Barran, P. E., Smerdon, S. J. & Carvalho, L. P. S. d. Uncoupling conformational states from activity in an allosteric enzyme. *Nat Commun* **8,** 203 (2017).

192. Moggré, G.-J., Poulin, M. B., Tyler, P. C., Schramm, V. L. & Parker, E. J. Transition State Analysis of Adenosine Triphosphate Phosphoribosyltransferase. *ACS Chem Biol* **12,** 2662–2670 (2017).

193. Cho, Y., Sharma, V. & Sacchettini, J. C. Crystal structure of ATP phosphoribosyltransferase from Mycobacterium tuberculosis. *J Biol Chem* **278,** 8333–9 (2003).

194. Lohkamp, B., McDermott, G., Campbell, S. A., Coggins, J. R. & Lapthorn, A. J. The structure of Escherichia coli ATP-phosphoribosyltransferase: identification of substrate binding sites and mode of AMP inhibition. *J Mol Biol* **336,** 131–44 (2004).

195. Ruszkowski, M. Guarding the gateway to histidine biosynthesis in plants: Medicago truncatula ATP-phosphoribosyltransferase in relaxed and tense states. *Biochem J* **475,** 2681–2697 (2018).

196. Champagne, K. S., Sissler, M., Larrabee, Y., Doublié, S. & Francklyn, C. S. Activation of the hetero-octameric ATP phosphoribosyl transferase through subunit interface rearrangement by a tRNA synthetase paralog. *J Biol Chem* **280,** 34096–104 (2005).

197. Vega, M. C., Zou, P., Fernandez, F. J., Murphy, G. E., Sterner, R., Popov, A. & Wilmanns, M. Regulation of the hetero-octameric ATP phosphoribosyl transferase complex from Thermotoga maritima by a tRNA synthetase-like subunit. *Mol Microbiol* **55,** 675–86 (2005).

198. Bovee, M. L., Champagne, K. S., Demeler, B. & Francklyn, C. S. The quaternary structure of the HisZ-HisG N-1-(5'-phosphoribosyl)-ATP transferase from Lactococcus lactis. *Biochemistry* **41,** 11838–46 (2002).

199. Livingstone, E. K., Mittelstädt, G., Given, F. M. & Parker, E. J. Independent catalysis of the short form HisG from Lactococcus lactis. *FEBS Lett* **590,** 2603–10 (Aug. 2016).

200. Stroek, R., Ge, Y., Talbot, P. D., Glok, M. K., Bernaś, K. E., Thomson, C. M., Gould, E. R., Alphey, M. S., Liu, H., Florence, G. J., *et al.* Kinetics and Structure of a Cold-Adapted Hetero-Octameric ATP Phosphoribosyltransferase. *Biochemistry* **56,** 793–803 (Feb. 2017).

201. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67,** 271–81 (2011).

202. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69,** 1204–14 (2013).

203. Matthews, B. W. Solvent content of protein crystals. *J Mol Biol* **33,** 491–7 (1968).

204. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (Aug. 2007).

205. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50,** 760–3 (1994).

206. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62,** 1002–11 (2006).

207. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60,** 2126–32 (2004).

208. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53,** 240–55 (1997).

209. Frishman, D & Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **23,** 566–79 (1995).

210. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372,** 774–97 (2007).

211. Sigrist, C. J. A., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A. & Bucher, P. PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform* **3,** 265–74 (2002).

212. Sissler, M, Delorme, C, Bond, J, Ehrlich, S. D., Renault, P & Francklyn, C. An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis. *Proc Natl Acad Sci U S A* **96,** 8985–90 (1999).

213. Dietmann, S, Park, J, Notredame, C, Heger, A, Lappe, M & Holm, L. A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3. *Nucleic Acids Res* **29,** 55–7 (2001).

214. Feller, G. & Gerday, C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* **1,** 200–8 (2003).

215. Lonhienne, T, Gerday, C & Feller, G. Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. *Biochim Biophys Acta* **1543,** 1–10 (2000).

216. Low, P. S., Bada, J. L. & Somero, G. N. Temperature adaptation of enzymes: roles of the free energy, the enthalpy, and the entropy of activation. *Proc Natl Acad Sci U S A* **70,** 430–2 (1973).

217. Siddiqui, K. S. & Cavicchioli, R. Cold-adapted enzymes. *Annu Rev Biochem* **75,** 403–33 (2006).

218. Gerday, C. Psychrophily and catalysis. *Biology (Basel)* **2,** 719–41 (2013).

219. Isaksen, G. V., Åqvist, J. & Brandsdal, B. O. Protein surface softness is the origin of enzyme cold-adaptation of trypsin. *PLoS Comput Biol* **10,** e1003813 (2014).

220. Isaksen, G. V., Åqvist, J. & Brandsdal, B. O. Enzyme surface rigidity tunes the temperature dependence of catalytic rates. *Proc Natl Acad Sci U S A* **113,** 7822–7 (2016).

221. Alphey, M. S., Fisher, G., Ge, Y., Gould, E. R., Machado, T. F. G., Liu, H., Florence, G. J., Naismith, J. H. & da Silva, R. G. Catalytic and Anticatalytic Snapshots of a Short-Form ATP Phosphoribosyltransferase. *ACS Catalysis* **8,** 5601–5610 (2018).

222. Fisher, G., Thomson, C. M., Stroek, R., Czekster, C. M., Hirschi, J. S. & da Silva, R. G. Allosteric Activation Shifts the Rate-Limiting Step in a Short-Form ATP Phosphoribosyltransferase. *Biochemistry* **57,** 4357–4367 (2018).

223. Mittelstädt, G., Jiao, W., Livingstone, E. K., Moggré, G.-J., Nazmi, A. R. & Parker, E. J. A dimeric catalytic core relates the short and long forms of ATP-phosphoribosyltransferase. *Biochem J* **475,** 247–260 (2018).

Appendices

6.6 Appendix A

Purification Buffers

Heterocyclases	(MicD, PatD, LynD, OscD, TenD and fusion enzymes)
Lysis buffer (A)	20 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 3 mM BME
Elution buffer (B)	20 mM Tris pH 8.0, 500 mM NaCl, 250 mM imidazole, 3 mM BME
Low salt buffer	20 mM Tris pH 8.0, 100 mM NaCl, 3 mM BME
High salt buffer	20 mM Tris pH 8.0, 1 M NaCl, 3 mM BME
Gel filtration buffer	10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP

PatE	
Lysis buffer	Lysis buffer (A) + 8M Urea
Elution buffer	Elution buffer (B) + 8M Urea
Gel filtration buffer	10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP

Oxidase	(ArtGox)
Lysis buffer	Lysis buffer (A) + 50 µM FMN
Elution buffer	Elution buffer (B) + 50 µM FMN
Gel filtration buffer	10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP

Purification Buffers (continued)

*Pa*ATPPRT	(HisG and HisZ subunits)
Lysis buffer	50 mM HEPES pH 8.0, 10 mM imidazole and 500 mM NaCl
Elution buffer	50 mM HEPES pH 8.0, 500 mM imidazole and 500 mM NaCl
TEV-proteolysis buffer	20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM DTT, 10% Glycerol
Storage buffer	20 mM HEPES pH 8.0

*Mt*PPase	
Lysis buffer	20 mM HEPES pH 8.0, 20 mM imidazole, 500 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$
Elution buffer	20 mM HEPES pH 8.0, 250 mM imidazole, 500 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$
Gel filtration buffer	20 mM HEPES pH 8.0

6.7 Appendix B - Supplementary Figures

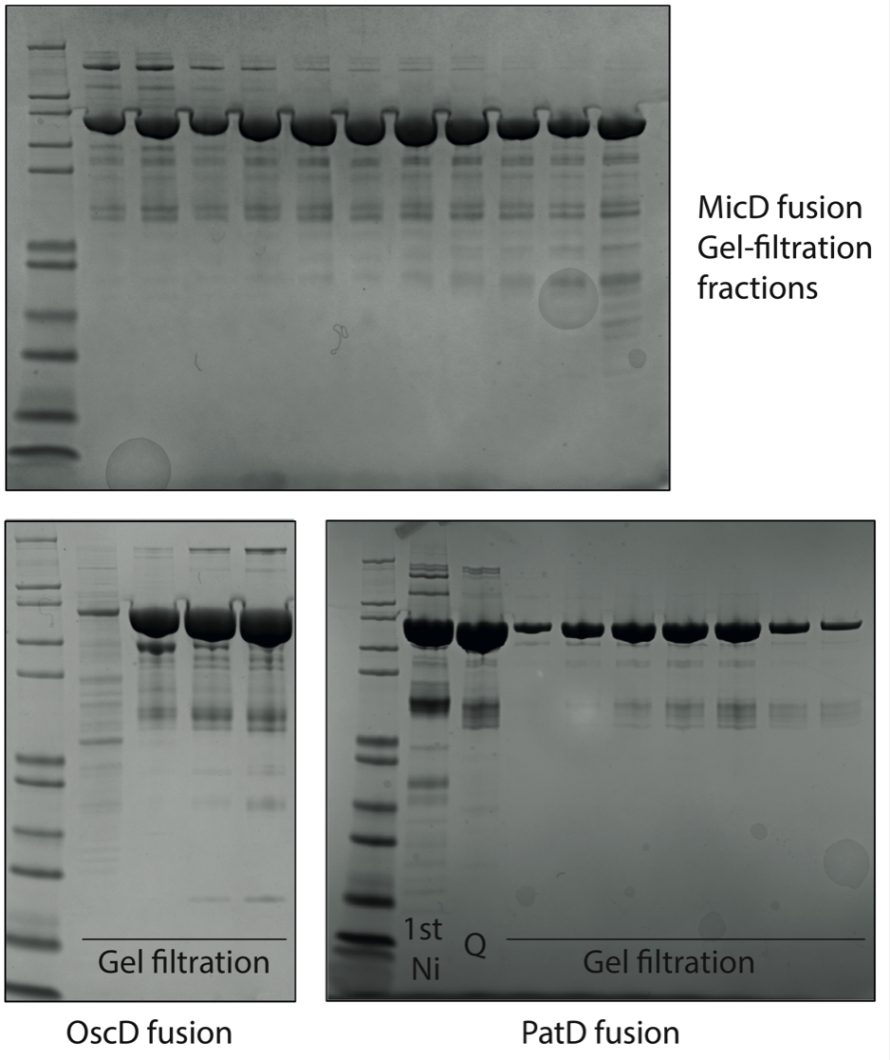

Figure S1: Uncropped SDS-PAGE images of MicD fusion, OscD fusion and PatD fusion following purification. Supplementary to Figure 2.4.

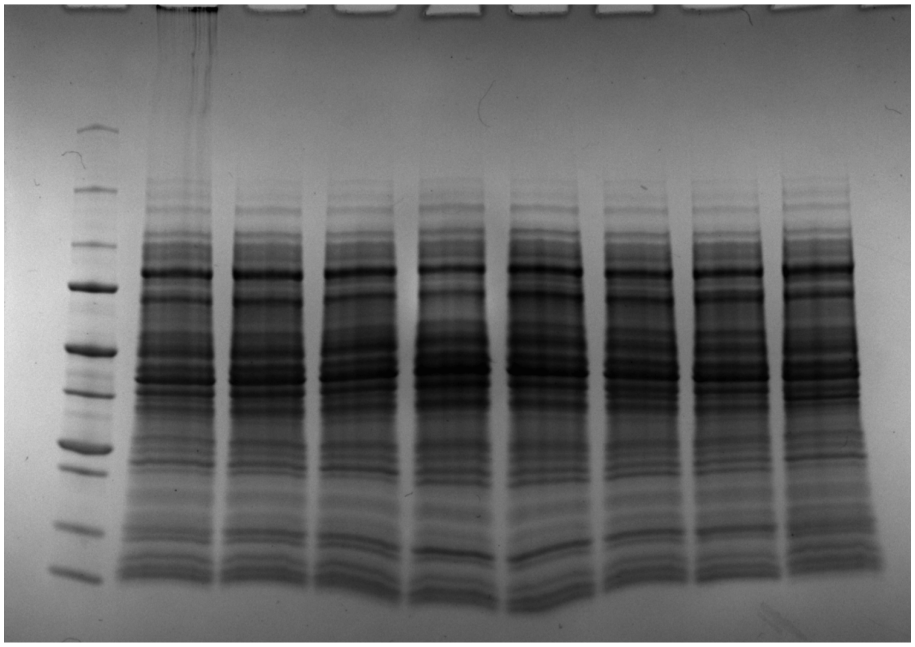

Figure S2: Original SDS-PAGE of MicD fusion expression test described in Figure 2.7A.

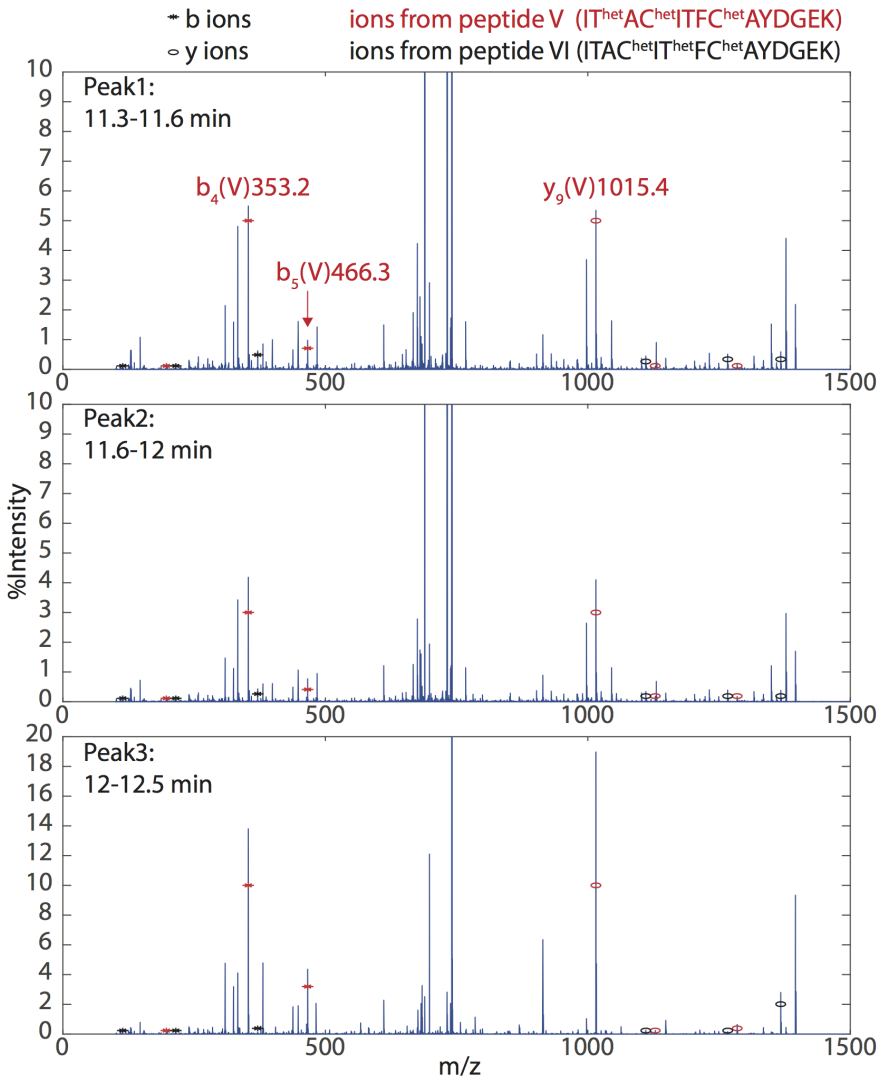

Figure S3: MSMS Spectra corresponding to peaks 1, 2 and 3 from Figure 3.14A.

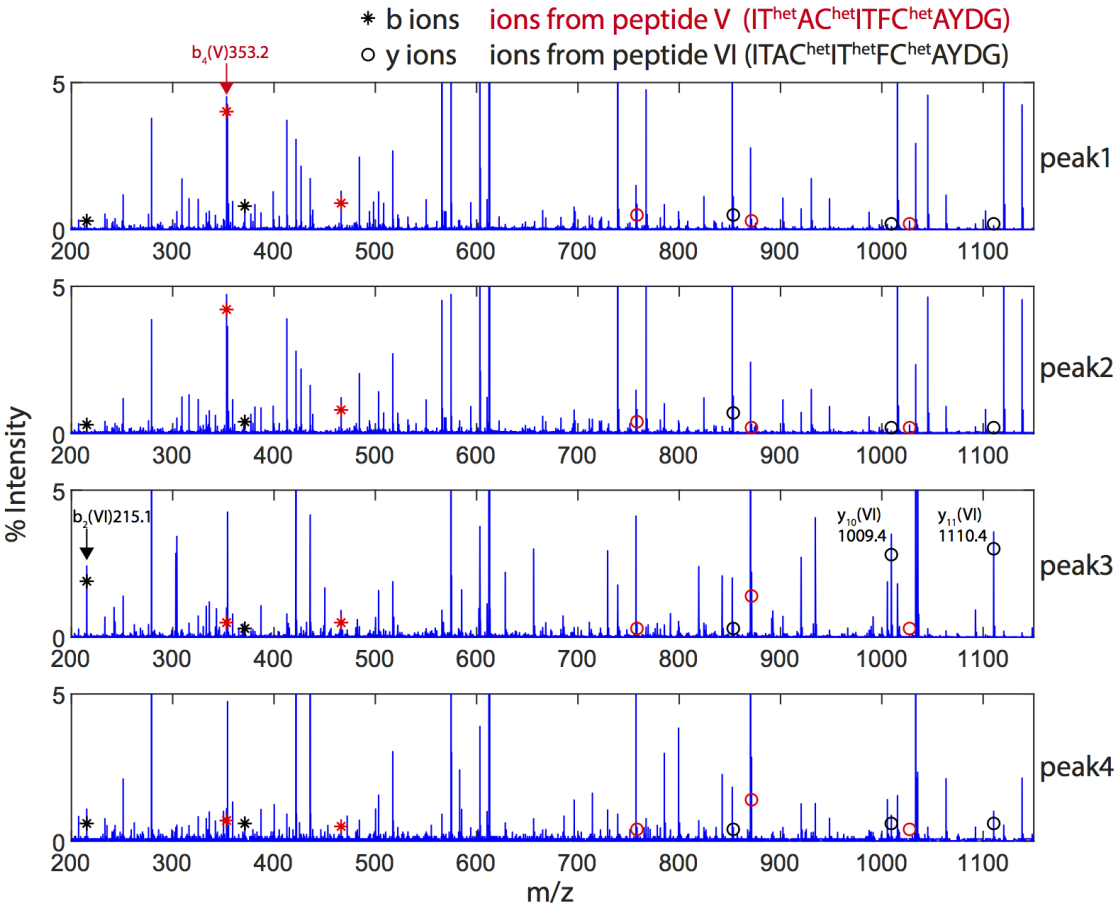

Figure S4: MSMS Spectra corresponding to peaks 1, 2, 4 and 4 from Figure 3.17A.

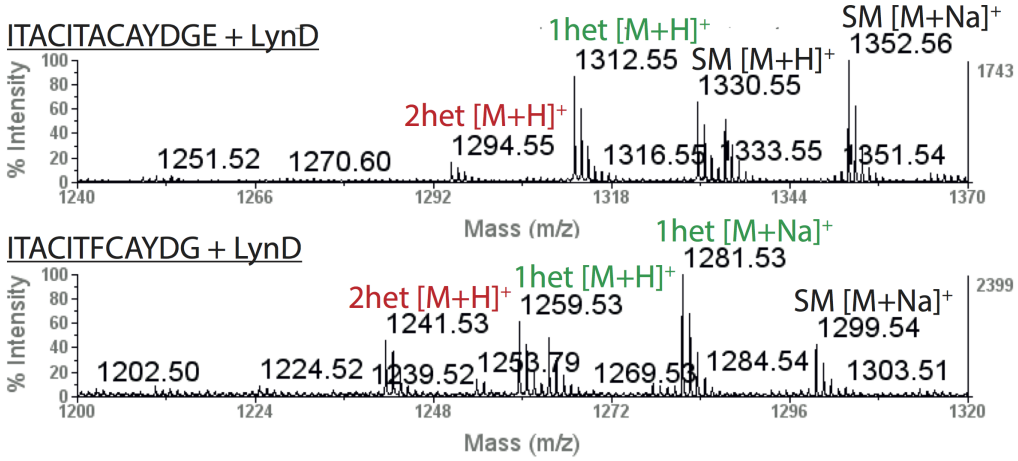

Figure S5: MALDI-MS Spectra of Leaderless peptides ITACITACAYDGE and ITACITFCAYDG reacted with native LynD (5 µM) for 16 h.